



Le processus de domiciliation des punaises hématophages vectrices de la maladie de Chagas : apport de l'étude du transcriptome chimiosensoriel

Axelle Marchant

► To cite this version:

Axelle Marchant. Le processus de domiciliation des punaises hématophages vectrices de la maladie de Chagas : apport de l'étude du transcriptome chimiosensoriel. Evolution [q-bio.PE]. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLS008 . tel-01350868

HAL Id: tel-01350868

<https://theses.hal.science/tel-01350868>

Submitted on 2 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLS008

**THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud**

**ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème**

Biologie

Par

Madame Axelle MARCHANT

**Le processus de domiciliation des punaises hématophages vectrices
de la maladie de Chagas : apport de l'étude du transcriptome chimiosensoriel**

Thèse présentée et soutenue à Gif-Sur-Yvette, le 15 janvier 2016 :

Composition du Jury :

M Claudio LAZZARI - Professeur (IRBI, Université François Rabelais) - Rapporteur
M Jean-Christophe SIMON - Directeur de Recherche (IGEPP, INRA) - Rapporteur
Mme Emmanuelle JACQUIN-JOLY - Directrice de Recherche (IEES, INRA) - Examinatrice
Mme Carole SMADJA - Chargée de Recherche (ISEM, CNRS) - Examinatrice
Mme Marie-Laure MARTIN-MAGNIETTE - Directrice de Recherche (IPS2, INRA Paris-Sud)
- Présidente du Jury
Mme Myriam HARRY - Professeur (Université Paris-Sud) - Directrice de thèse
Mme Florence MOUGEL - Maître de Conférence (Université Paris-Sud) - Co-encadrante



Sur mes cahiers d'écolier
Sur mon pupitre et les arbres
Sur le sable sur la neige
J'écris ton nom

Sur toutes les pages lues
Sur toutes les pages blanches
Pierre sang papier ou cendre
J'écris ton nom

Sur les images dorées
Sur les armes des guerriers
Sur la couronne des rois
J'écris ton nom

Sur la jungle et le désert
Sur les nids sur les genêts
Sur l'écho de mon enfance
J'écris ton nom

Sur les merveilles des nuits
Sur le pain blanc des journées
Sur les saisons fiancées
J'écris ton nom

Sur tous mes chiffons d'azur
Sur l'étang soleil moisi
Sur le lac lune vivante
J'écris ton nom

Sur les champs sur l'horizon
Sur les ailes des oiseaux
Et sur le moulin des ombres
J'écris ton nom

Sur chaque bouffée d'aurore
Sur la mer sur les bateaux
Sur la montagne démente
J'écris ton nom

Sur la mousse des nuages
Sur les sueurs de l'orage
Sur la pluie épaisse et fade
J'écris ton nom

Sur les formes scintillantes
Sur les cloches des couleurs
Sur la vérité physique
J'écris ton nom

Sur les sentiers éveillés
Sur les routes déployées
Sur les places qui débordent
J'écris ton nom

Sur la lampe qui s'allume
Sur la lampe qui s'éteint
Sur mes maisons réunies
J'écris ton nom

Sur le fruit coupé en deux
Du miroir et de ma chambre
Sur mon lit coquille vide
J'écris ton nom

Sur mon chien gourmand et
tendre
Sur ses oreilles dressées
Sur sa patte maladroite
J'écris ton nom

Sur le tremplin de ma porte
Sur les objets familiers
Sur le flot du feu béni
J'écris ton nom

Sur toute chair accordée
Sur le front de mes amis
Sur chaque main qui se tend
J'écris ton nom

Sur la vitre des surprises
Sur les lèvres attentives
Bien au-dessus du silence
J'écris ton nom

Sur mes refuges détruits
Sur mes phares écroulés
Sur les murs de mon ennui
J'écris ton nom

Sur l'absence sans désir
Sur la solitude nue
Sur les marches de la mort
J'écris ton nom

Sur la santé revenue
Sur le risque disparu
Sur l'espoir sans souvenir
J'écris ton nom

Et par le pouvoir d'un mot
Je recommence ma vie
Je suis né pour te connaître
Pour te nommer

Liberté.

Paul Eluard

Remerciements

Voilà que s'achèvent déjà ces trois années de thèse... Ces années auront été riches en apprentissage. J'ai acquis de nombreuses connaissances scientifiques et méthodologiques, mais aussi une certaine compréhension du monde à la fois passionnant et parsemé d'embûches qu'est celui de la recherche. Ce projet m'a également permis de travailler sur moi-même : curiosité, patience, remise en question et persévérance font aujourd'hui partie de moi. Ce projet n'aurait jamais pu aboutir sans toutes les personnes qui m'ont soutenue et qui se sont investies à mes côtés.

Je tiens tout d'abord à remercier les membres du jury, Claudio Lazzari, Jean-Christophe Simon, Carole Smadja, Marie-Laure Martin-Magnette et Emmanuelle Jacquin-Joly qui ont accepté de juger ce travail et d'apporter leur vision extérieure.

Je tiens également à remercier tous nos collaborateurs qui ont participé à la collecte des échantillons. Sans eux, ce projet n'aurait jamais pu aboutir. Merci à Carlos Eduardo Almeida et Jane Costa et leur équipe pour l'échantillonnage sur le terrain de l'espèce *Triatoma brasiliensis*. Merci à Vagner Mendonça, João Aristeu da Rosa, Marion Quartier et Patrick M. Guerin pour nous avoir fourni les souches de *Rhodnius*. Et merci à Jean-Michel Béranger et Denis Blanchet pour leur échantillonnage sur le terrain en Guyane de *Rhodnius robustus*.

Merci également à toutes les personnes qui nous ont conseillées, aidées, orientées dans ce travail. Ainsi, je tiens à remercier les membres du comité de thèses: Stéphan Le Crom, Fabrice Legei, Emmanuelle Jacquin-Joly et Cécile Neuveglise ; ainsi que Rachel Legendre, Claire Toffano et Andréa Rau.

Un très grand merci à ma directrice de thèse, Myriam. Merci pour m'avoir remis ce projet avec confiance. Merci de m'avoir toujours poussé à mettre en valeur mes résultats, par la publication ou par la participation à de nombreuses conférences et de m'avoir aidé à braver ma peur de parler en public en anglais. Merci pour ton soutiens scientifique, académique et humain. Merci également de m'avoir soutenue dans les recherches de financement tout au long de ce projet. Merci encore pour ces moments d'échanges informels et humains que nous avons partagé lors des pauses café, au restaurant lors d'un congrès...

Florence, je tiens tout particulièrement à te remercier. Ton aide m'a été très précieuse, tant d'un point de vue technique qu'analytique. Tant de questionnements et d'obstacles bioinformatiques franchis ensemble ! Et toujours dans l'humour et la bonne humeur ! Je me souviens de mes débuts dans ce vaste monde de Linux et du RNAseq, j'étais totalement perdue... Et M^{lle} Florence est arrivée à ma rescousse ! Tu m'as guidée à travers ces chemins complexes, tu m'as redonné confiance en moi pour enfin pouvoir être autonome mais toujours à l'écoute de tes précieux conseils. Merci aussi pour ce petit resto du 11 novembre qui a bien égayé ce jour férié passé à rédiger.

Merci à Pierre et à Catherine qui m'ont accueillie dans leur chaleureux laboratoire et merci à Jean-François et à Laure qui m'ont reçue dans l'équipe DEEIT. Je me suis très vite sentie à l'aise dans ce lieu de travail où règne une atmosphère humaine et de passion pour la recherche. Laure, merci également pour nos covoiturages à mes débuts, c'était très

sympathique. Dommage que mon emploi du temps n'ait pas permis de poursuivre cette habitude écologique et conviviale. Merci également pour avoir sauvé ma valise égarée à Rennes un jour de congrès... Merci également à Annie, Sylvie et Hélène pour leur aide dans toutes les démarches administratives. Merci à Nicolas d'avoir passé quelques longues soirées à m'aider dans mes lignes de codes. Merci à Jonathan, qui a eu la patience de faire tourner plusieurs fois le script de recherche de gènes d'intérêt et qui a toujours pris le temps de répondre à mes questions. Merci également à Elsa, qui m'a fourni de précieux conseils pour mes futures recherches de post-doctorats. Merci à tous les membres du labo et de l'équipe pour m'avoir aussi bien accueillie.

Je tiens également à remercier Claire, la reine de la BM qui m'a beaucoup aidée dans bien des manips. Mais surtout, merci Claire pour toutes nos conversations, tant scientifiques qu'humaines, qui m'ont beaucoup aidée dans les moments difficiles et m'ont bien divertie dans nos moments d'euphorie. Et ce n'est pas fini, bientôt Massive Attack !

Merci également à François qui a partagé mon bureau pendant une grande partie de ma thèse. Merci François pour toutes les fois où tu es venu m'aider dans des problématiques informatiques. Merci également pour tous tes conseils « politiques » et scientifiques. Ils m'ont fait progresser et je les applique au quotidien. Merci également pour nos conversations plus informelles sur la vie de tous les jours, l'expatriation et pour avoir partagé ta joie d'être papa. J'espère qu'on aura l'occasion de se voir en Bolivie.

Merci à tous les doctorants et « non-permanants » pour nos « réunions », pour toutes les bêtises que nous avons pu raconter et pour avoir refait le monde ensemble : merci Florian, Antoine, Pierre, Bastien, Quentin, Julien, Arnaud, Estelle et tous les autres. Un petit mot particulier pour Hanna : merci pour le thé, il a réchauffé et adouci mes journées de rédaction ! Merci aussi à Céline qui m'a tenue compagnie de nombreuses fois sur mes trajets du matin ! Une pensée aussi pour les anciens, en particulier Floriane et Morgane (¿Cuándo vamos a bailar salsa?). Merci à Seynabou et Estephania, doctorantes venues de l'autre bout du monde qui ont partagé mon bureau pendant quelques mois et avec qui j'ai pu échanger des conversations passionnantes. Merci également à tous les stagiaires qui sont passés pendant ces trois années de stage et qui ont apporté un peu de fraîcheur. Merci tout particulièrement à Romain et à Jérôme avec qui j'ai travaillé au quotidien pendant quelques mois.

Merci à tous mes amis d'avoir su continuer à me faire vivre et rire pendant ces trois ans de travail intense. Merci à Lucile, Cora, Nini de m'avoir soutenue dans mes moments de doutes et de fatigue.

Merci à mes parents et à ma grand-mère, qui ont toujours cru en moi, qui m'ont toujours donné la possibilité d'aller au bout de mes ambitions, tout en accordant une très grande importance à mon bonheur et à ma liberté. Un merci infini pour l'amour que vous me donnez, il est mon pilier, ma force et m'aide à avancer chaque jour.

Et enfin, merci à toi Rémi de m'avoir soutenue tous les jours et d'avoir cru en moi. C'est un plaisir de partager avec toi notre passion pour la photographie, de partir ensemble à la découverte de nouveaux paysages, de nouvelles cultures, de nouveaux mets culinaires... Merci de m'accorder chaque jour des moments de tendresse, d'amour et de rire. Je suis très heureuse de partager tous ces moments de bonheur avec toi et j'ai hâte de partir ensemble, à l'aventure !

Résumé

En Amérique Latine, les punaises hématophages Triatominae transmettent à l'homme le parasite *Trypanosoma cruzi*, responsable de la maladie de Chagas touchant actuellement 5 millions de personnes. Même si les programmes d'éradication chimique des vecteurs sont efficaces, la maladie persiste du fait de la recolonisation des habitations humaines par des vecteurs provenant d'habitats naturels. Ainsi, certaines espèces présentent une capacité d'adaptation aux anthrosystèmes (processus de domiciliation), alors que d'autres espèces apparentées ne l'ont pas. Comprendre cette capacité d'adaptation est crucial d'un point de vue épidémiologique afin de cibler les espèces présentant un risque pour l'homme. La capacité à s'adapter à un nouvel habitat pourrait être liée à l'évolution du répertoire de gènes du système chimiosensoriel, important pour la perception du milieu.

Cette étude a porté sur le système chimiosensoriel des Triatominae dans le but de documenter le processus d'adaptation et donc de domiciliation des vecteurs. Des données transcriptomiques obtenues en séquençage à haut débit ont été utilisées pour annoter et répertorier les gènes chimiosensoriels ainsi que pour comparer leur expression au sein de punaises hématophages d'habitats différents. L'existence d'une relation entre les variations de ces gènes chez différentes espèces de Triatominae et leur capacité d'adaptation à un habitat a par la suite été évaluée.

L'espèce *T. brasiliensis* en voie de domiciliation au Brésil et présentant à la fois des populations sylvatiques, péri-domiciliaires et domiciliaires, et différentes espèces du genre *Rhodnius* d'habitats variés, ont été étudiées, notamment les deux espèces sœurs, *R. robustus*, sylvatique en Amazonie et *R. prolixus* majoritairement domiciliée dans toute son aire de répartition.

En l'absence de génomes de références suffisamment proches de *T. brasiliensis* et des 10 espèces de *Rhodnius* étudiées, leurs transcriptomes ont été assemblés *de novo*. Les transcriptomes des deux espèces *R. prolixus* et *R. robustus* ont été assemblés par alignement sur le génome de *R. prolixus*. Chez ces différentes espèces de Triatominae étudiées, l'analyse du répertoire des gènes chimiosensoriels codant les OBPs et CSPs (familles multigéniques) comparé à celui d'autres Paraneoptères a montré des expansions géniques pouvant refléter des processus adaptatifs. Par ailleurs, chez les différentes espèces du genre *Rhodnius*, il existe une corrélation positive entre le nombre de gènes codant les OBPs et la capacité de domiciliation, suggérant l'implication de cette famille de gènes dans l'adaptation au milieu anthropique.

Les analyses d'expression différentielle concernant les différentes populations de *T. brasiliensis* et les espèces *R. prolixus*/*R. robustus* ont montré qu'un certain nombre de transcrits sont différentiellement exprimés selon l'environnement dans lequel ont évolué les punaises notamment des gènes chimiosensoriels (OBPs, CSPs) ainsi que des gènes impliqués dans le rythme circadien et le comportement de recherche alimentaire (*takeout*), dans la

réponse à des stress environnementaux comme des gènes de détoxification (P450, glutathione S-transférase), dans la résistance à des changements climatiques (heat-shock protéines) et dans la protection du milieu extérieur (protéines cuticulaires).

Ce travail a permis de mettre à la disposition de la communauté scientifique des outils performants pour l'étude du processus de domiciliation des vecteurs de la maladie de Chagas (transcriptome, répertoire de gènes). Il a également permis de révéler des gènes qui pourraient être impliqués dans l'adaptation et/ou la plasticité phénotypique en réponse à un changement d'habitat. La compréhension des bases moléculaires de l'adaptation des vecteurs aux habitations humaines ouvre des potentialités de développer des méthodes alternatives de lutte contre les vecteurs qui pourraient être basées sur une perturbation de la communication chimique.

Abstract

In Latin America, the bloodsucking bugs (Triatominae, Hemiptera, Reduviidae) are vectors of the parasite *Trypanosoma cruzi*, which causes Chagas disease. More than five million people are infected. Even if chemical control campaigns are effective against vectors, the disease persists due to the recolonization of human habitations by vectors from natural habitats. Some species have the capacity to adapt to anthroposystems (domiciliation process), while other related species do not. Understanding this capacity to adapt is crucial from an epidemiological perspective to target species at risk to humans. The capacity to adapt to a new habitat could be linked to changes in the repertoire of chemosensory system genes, particularly for odorant binding proteins (OBP) and chemosensory proteins (CSP), which are important proteins to detect various odor stimuli.

This study is based on the chemosensory system of Triatominae to document the adaptation process and then the domiciliation of the vectors. Transcriptomic data obtained by high-throughput sequencing were used to annotate and list the chemosensory genes and also to compare their expression in bloodsucking bugs from different habitats. The relationship between changes in these genes in different Triatominae species and their ability to adapt to a new habitat was evaluated.

The species *T. brasiliensis*, which is in the process of domiciliation in Brazil with sylvatic, peridomestic and domestic populations, and various species of the genus *Rhodnius* from diverse habitats were studied, especially the two sibling species *R. robustus*, sylvatic in the Amazonia and *R. prolixus* mostly domestic throughout its geographical range.

In the absence of a reference genome for *T. brasiliensis*, a reference transcriptome *via de novo* assembly (data 454 and Illumina) was achieved. The reference transcriptomes for 10 *Rhodnius* species were also established using the *de novo* assembly method. A genome reference based method on *R. prolixus* was also used to assemble the transcriptome of the two species *R. prolixus* and *R. robustus*. In the different species of the Triatominae studied, the chemosensory gene repertoire showed a high diversity and genic expansions compared to that of others Paraneoptera, which could reflect adaptive process. Furthermore, a positive correlation was shown between the number of OBP genes in *Rhodnius* species and their domiciliation ability, suggesting that this gene family is involved in the adaptation to anthropogenic environment.

The differential expression analyses on the *T. brasiliensis* populations and the *R. prolixus* / *R. robustus* species showed that some transcripts are differentially expressed according to the environment in which the bugs have evolved, especially the chemosensory genes (OBP, CSP) and also genes involved in the circadian rhythm and foraging behavior (*takeout*), in the response to environmental stress such as detoxification genes (P450, glutathione S-transferase), in resistance to climatic changes (heat-shock proteins) and in protection from the external environment (cuticular proteins).

This work has helped make available to the scientific community powerful tools for studying the process of domiciliation of Chagas disease vectors (transcriptome, gene repertoire). It also revealed genes that could be involved in the adaptation and/or phenotypic plasticity in response to a change in habitat.

Understanding the molecular basis of vector adaptation to human dwellings opens the potential to develop new tools to control the disease vectors, for example by disrupting chemical communication.

Table des matières

Remerciements	
Résumé.....	
Abstract	
Problématique Générale	1
Objectifs.....	3
Plan de la thèse.....	4
Introduction	5
I. La maladie de Chagas	7
1. Cycle épidémiologique	7
2. Pathologie, diagnostic et traitement de la maladie.....	8
3. Distribution de la maladie.....	9
4. Les vecteurs de la maladie de Chagas	10
5. Cas des espèces du genre <i>Rhodnius</i> et du complexe d'espèces <i>T. brasiliensis</i>	13
II. La maladie de Chagas dans le contexte des changements globaux.....	16
1. Risque d'émergence	16
2. La lutte antivectorielle et le processus de domiciliation	18
III. L'apport du séquençage à haut débit en biologie de l'évolution.....	20
1. Séquençage complet des génomes	20
2. Le cas du RNAseq.....	21
3. Chez les Triatominae.....	25
IV. Le système chimiosensoriel chez les insectes.....	26
1. Les molécules semiochimiques	26
2. Mode de fonctionnement du système chimiosensoriel.....	26
3. Le système chimiosensoriel chez les Triatominae	32
Matériels et Méthodes	35
I. Echantillonnage.....	37
1. Triatoma brasiliensis	37
2. Les 10 espèces du genre <i>Rhodnius</i>	37
II. Analyse de données RNAseq	38
1. Préparation des données.....	38
2. Alignement, comptage et analyse statistique.....	41
3. Corrélation entre le nombre de CSPs et OBPs définies chez les <i>Rhodnius</i> et leur capacité de domiciliation.....	42
Partie I : Analyse du transcriptome chimiosensoriel de <i>Triatoma brasiliensis</i>.....	43
Chapitre 1 : Elaboration et annotation d'un transcriptome de référence.....	45
Article 1: <i>De novo</i> transcriptome assembly for a non-model species, the blood-sucking bug <i>Triatoma brasiliensis</i> , a vector of Chagas disease (Genetica, April 2015, Volume 143, Issue 2, pp 225-239).	47
Chapitre 2 : Analyse de l'expression différentielle entre populations de <i>T. brasiliensis</i> provenant d'environnements différents	73
Article 2: Under expression of chemosensory genes in domiciliary bugs of the Chagas disease vector <i>Triatoma brasiliensis</i>	75

Partie II : Analyse des transcriptomes chimiosensoriels chez différentes espèces du genre <i>Rhodnius</i>	109
Chapitre 1 : Elaboration et annotation du transcriptome de <i>Rhodnius prolixus</i>	111
Article 3: Comparing de novo and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug <i>Rhodnius prolixus</i> (Article in press in Insect Biochemistry and Molecular Biology 2015).....	113
Chapitre 2 : Analyse d'expression des gènes chimiosensoriels entre populations, sexes et tissus, de deux espèces sœurs de <i>Rhodnius</i> , <i>R. prolixus</i> et <i>R. robustus</i>	133
Chapitre 3 : Evolution du nombre de gènes chimiosensoriels : lien avec la capacité de domiciliation.....	173
Discussion et Perspectives	175
Discussion.....	177
Limites de l'étude.....	177
Utilisation du modèle Triatominae pour l'évaluation de différentes méthodes d'assemblage RNAseq et d'analyse d'expression.....	179
Implication de l'expression des transcrits des organes chimiosensoriels dans l'adaptation aux anthroposystèmes chez les Triatominae	183
Plasticité phénotypique ou adaptation génique ?.....	186
Evolution des répertoires de gènes chimiosensoriels chez les Triatominae	188
Perspectives.....	192
Elaboration d'un nouveau plan d'échantillonnage.....	192
Mécanismes adaptatifs	192
Etudes fonctionnelles des gènes chimiosensoriels.....	193
Utilisation des gènes chimiosensoriels dans la lutte vectorielle	194
Communications.....	196
Annexes.....	197
Références	201

Problématique Générale

Les maladies vectorielles impliquant le plus souvent un arthropode (insecte ou acarien) hématophage sont responsables de plus de 17% des maladies infectieuses et provoquent plus d'un million de décès chaque année selon l'OMS (Mars 2014). Parmi elles, le paludisme responsable de plus de 600 000 décès par an, est probablement la plus connue. Bien d'autres maladies vectorielles d'importance majeure sont répertoriées par l'OMS : Dengue, Chikungunya, fièvre jaune, encéphalite japonaise, Filariose lymphatique (éléphantiasis), leishmaniose, fièvre hémorragique de Crimée-Congo, maladie de Chagas, maladie de Lyme, Schistosomiase, Trypanosomiase humaine africaine (maladie du sommeil)... Sur les 17 maladies tropicales négligées, 6 sont des maladies vectorielles. C'est le cas de la Dengue, de la maladie de Chagas, de la Trypanosomiase humaine africaine, de la Leishmaniose, de la Filariose lymphatique, de l'Onchocercose, de la Schistosomiase. Ces maladies tropicales négligées sont fréquentes dans les zones rurales isolées ou dans les bidonvilles d'Afrique, d'Asie et d'Amérique au sein des populations à faible revenu qui n'ont pas accès aux éventuels vaccins ou traitements. Les personnes touchées ou qui sont exposées à ces maladies ont peu de possibilités d'expression sur le terrain politique du fait de leur exclusion sociale ou géographique. De ce fait, on entend peu parler des maladies tropicales négligées. Cependant, elles provoquent souffrances et décès et nécessiteraient des campagnes de prévention et de traitement importantes (OMS, 2011). Parmi elles, nous nous intéresserons plus particulièrement à la Maladie de Chagas. Cette maladie potentiellement mortelle est provoquée par un trypanosome (*Trypanosoma cruzi*). Ce parasite est transmis d'un vertébré à un autre par les punaises hématophages Triatominae. La maladie de Chagas est distribuée dans toute l'Amérique Latine et Centrale. Des stratégies de contrôle anti-vectoriel ont été mises en place à des échelles nationales et internationales sur tout le continent afin de lutter contre la maladie. Ainsi, les deux principaux vecteurs domiciliaires, *Triatoma infestans* et *Rhodnius prolixus*, ont été éliminés de certains pays. Ces initiatives ont permis de diminuer considérablement la prévalence de la maladie. Ainsi, de 18 millions de personnes infectées dans les années 90, l'OMS en compte aujourd'hui 5 millions. Cependant, l'élimination des espèces principales de vecteurs ouvre de nouvelles niches écologiques à des espèces jusque là sylvatiques. Ainsi, la colonisation des habitats humains par de nouvelles espèces est observée. Ce phénomène est appelé « processus de domiciliation ». Après l'éradication des populations domiciliées de *T. infestans* au Brésil, *Triatoma brasiliensis* est devenu le vecteur principal de la maladie de Chagas au nord du Brésil. Certaines espèces de Triatominae présentent cette capacité d'adaptation, de domiciliation, comme *T. brasiliensis*, alors que d'autres espèces apparentées ne l'ont pas. Par exemple, les deux espèces sœurs du genre *Rhodnius*, *R. prolixus* et *R. robustus* présentent une

divergence d'habitat : *R. robustus* est une espèce sylvatique, inféodée aux palmiers, alors que *R. prolixus* est domiciliée.

L'apparition de changements majeurs dans l'environnement à l'échelle du globe va impacter fortement le risque d'émergence ou de ré-émergence des maladies vectorielles. Parmi ces changements globaux, le climat a un rôle très important dans l'extension de l'aire de distribution des pathogènes ou des vecteurs. Par ailleurs, les activités humaines (l'urbanisation, le développement industriel, la mondialisation des échanges, la déforestation, l'intensification des mouvements d'animaux ...) accélèrent considérablement l'évolution de l'environnement.

La modification des écotypes naturels des punaises par des actions anthropiques peut également intervenir dans le changement de répartition des vecteurs de la maladie de Chagas. En effet, face à l'anthropisation de leur habitat, les espèces répondent en se déplaçant ou en évoluant ; en d'autres termes, en s'adaptant. La déforestation au profit d'implantation de cultures intensives pourrait conduire différentes espèces de Triatominae à quitter leur habitat naturel et à coloniser d'avantage les milieux anthropiques.

Objectifs

Ce travail de thèse s'inscrit dans une problématique plus générale de la réponse adaptative des insectes aux anthroposystèmes qui a fait l'objet de l'ANR ADAPTANTHROP (2010-2013, M. Harry, resp). Parmi les modèles d'étude de ce projet, les vecteurs de la maladie de Chagas ont été utilisés afin d'appréhender le processus de domiciliation.

Dans ce contexte, notre problématique générale était de comprendre pourquoi certaines espèces, comme *R. prolixus*, sont totalement domiciliées, d'autres, comme *T. brasiliensis*, présentent des populations aussi bien sylvatiques que domiciliées, alors que d'autres encore, comme *R. robustus*, ne semblent pas capables de coloniser les milieux anthropisés.

L'objectif principal était de documenter la réponse adaptative des vecteurs de la maladie de Chagas en réponse à une pression anthropique.

En corollaire, les questions soulevées étaient les suivantes : Quels sont les gènes impliqués dans la domiciliation des vecteurs ? Est-ce que la réponse adaptative est contrôlée par le même jeu de gènes au sein des différentes espèces soumises à la pression anthropique ? Peut-on la mettre en évidence par des différences génétiques ou d'expression génique ? S'il y a adaptation, est ce plutôt de la plasticité phénotypique, ou l'action de la sélection sur des variants génétiques à l'origine de caractères avantageux ?

Par ailleurs, les insectes utilisant la communication chimique pour interagir avec leur environnement, le système chimiosensoriel a souvent été proposé comme acteur clé de la capacité d'un insecte à s'adapter à un nouvel habitat (Sánchez-Gracia et al., 2009). Nous nous sommes donc interrogés sur l'existence d'une relation entre les variations de ces gènes chez différentes espèces de vecteurs et leur capacité d'adaptation à un habitat.

Pour tenter de répondre à ces questions, nous avons choisi une approche transcriptomique utilisant le séquençage à haut débit avec une double approche :

- i) au niveau spécifique, nous avons utilisé une approche gènes candidats, en nous focalisant sur les gènes chimiosensoriels pour lesquels des études ont montré l'impact de la sélection, et
- ii) au niveau populationnel, nous avons adopté une approche de transcriptomique comparative en ciblant les organes chimiosensoriels, afin de caractériser des changements d'expression génique selon différentes conditions et d'identifier de nouveaux gènes ou groupes de gènes potentiellement impliqués dans l'adaptation.

Les modèles utilisés ont été d'une part i) le modèle *Triatoma brasiliensis*, qui possède des populations réparties dans les trois types d'environnement et ii) le modèle *Rhodnius* avec la comparaison de a) différents populations/souches des deux espèces sœurs, *R. prolixus* (domiciliée) et *R. robustus* (sylvatique), b) de dix espèces de *Rhodnius* documentant un gradient de domiciliation.

De plus, l'annotation de gènes chimiosensoriels et les analyses transcriptomiques comparatives ont nécessité l'assemblage de transcriptomes et la construction d'un transcriptome de référence

pour nos espèces modèles. Ceci nous a conduits à nous poser des questions méthodologiques sur les assemblages et sur l'analyse des données transcriptomiques.

Plan de la thèse

Le plan de ma thèse suit notre cheminement scientifique.

Dans un premier temps, j'apporterai en introduction des renseignements plus spécifiques en ce qui concerne la maladie de Chagas, le parasite qui en est la cause, et la biologie de ces vecteurs. Des informations sur le système chimiosensoriel des insectes et leur évolution seront apportées afin de mieux comprendre son implication dans les processus adaptatifs. Enfin, une rapide description des différentes méthodes de séquençage à haut débit sera présentée afin d'appréhender leurs intérêts et leurs problématiques techniques. Par la suite, la description des échantillons qui ont permis de réaliser cette étude ainsi que les différentes méthodes d'analyse de données RNAseq seront présentées dans la section « Matériels et Méthodes ».

Mes résultats seront exposés sous forme d'articles. Cependant, toutes les références bibliographiques seront regroupées à la fin du mémoire.

- La première partie concerne l'analyse du transcriptome chimiosensoriel de l'espèce *T. brasiliensis* sera exposée. Elle se divise en deux chapitres. Le premier chapitre a fait l'objet d'un article publié dans le journal *Genetica*. Cet article décrit l'élaboration et l'annotation du transcriptome de référence pour cette espèce. Le second chapitre constitue un article soumis dans le journal *Plos Neglected Tropical Diseases* et décrit l'analyse de l'expression différentielle entre populations de *T. brasiliensis* provenant d'environnements différents.

- La seconde partie concerne l'analyse des transcriptomes chimiosensoriels de différentes espèces du genre *Rhodnius*. Elle se divise en trois chapitres. Le premier a fait l'objet d'un article dans le journal *Insect Biochemistry and Molecular Biology* sur l'élaboration et l'annotation du transcriptome de *R. prolixus*. L'analyse de l'expression génique entre des populations des deux espèces sœurs de *R. prolixus* et *R. robustus* inféodées à des environnements différents sera présentée dans le deuxième chapitre. Dans le dernier chapitre, nous exposerons la comparaison des répertoires chimiosensoriels de différentes espèces du genre *Rhodnius*, et tenterons de la relier au gradient de domiciliation tout en tenant compte du signal phylogénétique.

Dans une discussion générale, nous comparerons les résultats po distincts, puis dans une perspective nous aborderons les ouvertures fondamentales et appliquées que permettent d'envisager notre étude.

Introduction

I. La maladie de Chagas

La maladie de Chagas, endémique à l'Amérique Latine et découverte par Carlos Chagas en 1909, est provoquée par le parasite *Trypanosoma cruzi* (Kinétoplastidae, Trypanosomatidae). Ce parasite est principalement transmis d'un hôte vertébré à un autre par l'intermédiaire de punaises hématophages vectrices : les Triatominae (Hemiptera, Reduviidae). Malgré le succès des campagnes d'éradication de certains vecteurs, la maladie de Chagas demeure une maladie tropicale négligée de grande incidence socio-économique en Amérique Latine. Ainsi, on estime que le coût économique global de la maladie de Chagas dépasse 7 milliards de dollars USD annuellement (Lee et al., 2013). Elle est considérée comme endémique dans 21 pays d'Amérique Latine où les vecteurs subsistent. L'OMS estime que 5 millions de personnes dans le monde seraient infectées (WHO, 2015) et que plus de 100 millions de personnes seraient exposées au risque de cette maladie. Chaque année, 300 000 nouveaux cas d'infection et 50 000 décès (dont 15 000 au Brésil) sont répertoriés.

1. Cycle épidémiologique

Le cycle de développement post embryonnaire chez des vecteurs Triatominae est composé de 5 stades juvéniles tous hématophages comme les adultes et donc susceptibles d'être infectés par *T. cruzi*. La durée de développement pour passer de l'œuf au stade adulte peut s'étendre sur 3 à 6 mois selon l'espèce mais aussi selon les conditions de température et d'humidité, la densité de population et la fréquences des repas sanguins (Gourbière et al., 2012; Luz et al., 1999; Schofield, 1979; Lazzari et al., 2013). La durée de vie des adultes peut aller de quelques mois (*R. prolixus*, *T. infestans*) à plus d'une année (*T. dimidiata*, *Panstrongylus megistus*).

Lors d'un repas sanguin sur un hôte infecté, les vecteurs Triatominae ingèrent les parasites. Le cycle de développement de *T. cruzi* comporte différents stades. (Figure 1). Les formes trypomastigotes ingérées par les vecteurs se transforment en épimastigotes qui se répliquent dans l'intestin de la punaise. Ces formes migrent dans l'intestin postérieur et se différencient en trypomastigotes métacycliques (dernier stade évolutif) dans le rectum. Lors d'un repas sanguin, les Triatominae défèquent sur la peau de leur hôte. Les trypomastigotes présents dans les fèces passent la barrière cutanée par la plaie causée par la piqûre ou pénètrent par les muqueuses. Les parasites se différencient dans les cellules de l'hôte en amastigotes, forme non flagellée du trypanosome. Après réplication puis lyse des cellules, les parasites sont libérés dans le sang sous forme de trypomastigotes. Sous cette forme flagellée, les parasites envahissent d'autres cellules pour poursuivre leur cycle ou être prélevés par un nouveau vecteur lors d'une piqûre (Rassi et al., 2010).

Figure 1 : voir figure 1 de Rassi, A., Rassi, A., and Marin-Neto, J.A. (2010). Chagas disease. The Lancet 375, 1388–1402. (<http://www.sciencedirect.com.gate1.inist.fr/science/article/pii/S014067361060061X?np=v#>)

T. cruzi aurait divergé de son dernier ancêtre commun avec *T. rangeli* il y a 23 à 39 millions d'années (Lewis et al., 2011). Il présente une diversité génétique considérable (Messenger et al., 2015). Le consensus international actuel reconnaît un minimum de six lignées génétiques stables (TcI-TcVI) ayant des répartitions géographiques distinctes (Zingales et al., 2009). Un septième génotype potentiel (TcBat) proche génétiquement de TcI, a récemment été rapporté dans le nord de l'Amérique du Sud (Marcili et al., 2009; Messenger et al., 2015; Pinto et al., 2012; Ramírez et al., 2014).

Hormis le type de transmission vectorielle décrit ci-dessus, il existe chez l'homme, un mode de contamination orale. L'ingestion de boissons ou de nourriture contaminées par des déjections de Triatominae sont associées à des infections massives par le parasite (Benchimol Barbosa, 2006; Pereira et al., 2009; Rassi et al., 2010). Cette contamination orale s'accompagne de taux de mortalité élevés (Barreto-de-Albuquerque et al., 2015; Sánchez and Ramírez, 2013). Notons également la possibilité de transmission du parasite d'homme à homme par transfusion sanguine ou transplantation d'organes en provenance de donneurs infectés ou encore d'une mère infectée à son enfant.

2. Pathologie, diagnostic et traitement de la maladie

La maladie de Chagas se développe en deux phases chez l'homme. La première phase appelée phase aigüe, dure environ deux mois. Cette phase peut être asymptomatique ou conduire à des accès fébriles ou encore à un œdème périorbitaire (signe de Romana, Figure 3) si la pénétration du parasite a lieu au niveau des muqueuses oculaires ou à une inflammation des ganglions lymphatiques). Une méningite aiguë mortelle peut être provoquée chez les enfants de moins de deux ans par la migration des parasites dans les méninges (Pentreath, 1995). Dix à vingt ans après une période d'infection « silencieuse », le parasite se multiplie à l'intérieur des cellules de l'hôte. La maladie peut alors évoluer en phase indéterminée (aucune pathologie apparente) ou en phase chronique (Figure 3) qui peut alors provoquer la mort. Cette dernière apparaît dans plus d'un tiers des cas. Selon les souches de parasites (et donc leur répartition géographique), différents organes peuvent être affectés (cœur, système digestif, système nerveux ; De Diego et al., 1991). Selon l'OMS jusqu'à 30% des patients souffrent des troubles cardiaques potentiellement mortels et jusqu'à 10% des patients présentent de troubles digestifs conduisant à une perte de poids ou/et des troubles neurologiques pouvant mener jusqu'à la démence (Prata, 2001; Rassi et al., 2010).

Les patients en phase indéterminée ne présentent pas de complications ni de symptômes. Cependant, ils jouent un rôle de réservoir pour les vecteurs et peuvent transmettre le parasite par transfusion sanguine (Quartier, 2011).

Au moment de la phase aigüe, il est possible de détecter la présence de *T. cruzi* dans le sang d'un patient infecté. La mise en évidence du parasite lors de la phase chronique nécessite

au préalable la multiplication du parasite par des procédures qui peuvent être longues à mettre en place : mise en culture du sang du patient ou xénodiagnostic qui consiste à nourrir des Triatominae par le sang du patient et à observer par examen au microscope 4 semaines plus tard, le contenu d'une goutte fécale des insectes nourris. Cependant, ce type d'examen peut conduire à des faux positifs, *T. cruzi* pouvant être confondu morphologiquement avec *T. rangeli* (trypanosome non pathogène pour les vertébrés et vivant en sympatrie avec *T. cruzi*). Il existe également des tests par diagnostic immunologique : détection d'anticorps, immuno-absorbance enzymatique, appelé plus couramment test ELISA, immunofluorescence indirecte ou hémagglutination indirecte (Lopez-Antunano et al., 2000; Rassi et al., 2010). Les patients atteints de leishmaniose, malaria, toxoplasmose ou d'infection bactérienne comme la lèpre ou la tuberculose peuvent cependant provoquer des faux positifs. C'est pourquoi deux tests positifs sont nécessaires pour confirmer le diagnostic (Teixeira et al., 2006). Plus récemment, la méthode PCR a été proposée pour le diagnostic de la maladie de Chagas en raison de sa grande sensibilité et de sa spécificité (Gilber et al., 2013).

Deux médicaments (benznidazole ou nifurtimox) sont efficaces pour lutter contre le parasite lors de la phase aigüe mais nécessitent un suivi médical. Coura et Castro, (2002) observent seulement 10 à 20% de succès des traitements durant la phase chronique. A l'heure actuelle, aucun vaccin efficace n'a été trouvé pour lutter contre la maladie de Chagas. C'est pourquoi la principale stratégie de prévention contre la maladie consiste à éradiquer chimiquement les vecteurs qui ont colonisé les habitations humaines et à contrôler des donneurs de sang ou d'organes (Dias et al., 2002; "OMS | Maladie de Chagas (trypanosomiase américaine)," n.d.). Des campagnes de lutte antivectorielle ont permis de diminuer considérablement le nombre de personnes infectées. Ainsi, l'Amérique latine est passée de plus de 17 millions de malades recensés en 1980 à 5 millions en 2015 (WHO, 2015). Cependant, il existe de nombreuses disparités entre les pays, notamment dues aux programmes de lutte antivectorielle mis en place (Coura and Viñas, 2010).

3. Distribution de la maladie.

La maladie de Chagas a longtemps été confinée à des régions rurales et pauvres d'Amérique Latine où le contact entre les hommes et les vecteurs étaient les plus fréquents. Toutefois, dans cette dernière décennie, la mobilité des populations entre l'Amérique Latine et le reste du monde l'a propagée aux Etats-Unis, au Canada, en Europe et dans certains pays du Pacifique occidental (Figure 2) (Schmunis, 2007; Schmunis and Yadon, 2010). En effet, les immigrants Latinos Américains ignorant qu'ils sont infectés peuvent transmettre l'infection à d'autres personnes par transfusion sanguine, don d'organes ou par transmission materno-fœtale. Aux Etats-Unis et en Espagne, le nombre de personnes infectées par *T. cruzi* s'élève à plus de 50 000 (Rassi et al., 2010). En France, la maladie de Chagas est présente de façon endémique en Guyane. En métropole, 9 cas importés y ont été diagnostiqués entre 2004 et 2006, (Lescure et al., 2008) et le risque de contamination par don de sang est estimé à 0,68 pour 100 000 par

an (Pillonel et al., 2009). Depuis 2007, l'établissement français du sang a mis en place le dépistage systématique de tous les donneurs de sang ayant séjourné en Amérique Latine (Lescure et al., 2008).

Figure 2 : Distribution de personnes infectées par *Trypanosoma cruzi* dans le monde en 2009 (voir http://thehealthcoach1.com/wp-content/uploads/2012/06/MapChagasJun09_large.jpg)

4. Les vecteurs de la maladie de Chagas

Les vecteurs de la maladie de Chagas sont des hémiptères appartenant à la famille des Reduviidae et à la sous-famille des Triatominae. Environ 150 espèces de Triatominae sont actuellement décrites et se répartissent en 15 genres (GALVÃO et al., 2003; Galvão and Justi, 2015; Schofield and Galvão, 2009). Les espèces de punaises hématophages peuvent pour la majorité d'entre elles être distinguées sur la base de caractères morphologiques. Cependant, certaines espèces présentent des caractères morphologiques très similaires et nécessitent d'autres outils pour les différencier. Inversement, certaines espèces présentent un polymorphisme intraspécifique très important (Gourbière et al., 2012). Afin de mieux distinguer les espèces, différents outils et tests ont été élaborés (cytogénétique, marqueurs moléculaires, croisements). A titre d'exemple, deux populations de *T. infestans* longtemps considérées comme deux espèces distinctes ont été regroupées dans une seule et même espèce après que leur croisement ait donné lieu à des descendants fertiles et qu'aucune différence n'ait été trouvée *via* des études cytogénétiques ou par marqueurs moléculaires (Bargues et al., 2006; Gourbière et al., 2012; Monteiro et al., 1999). Inversement, l'espèce *Triatoma brasiliensis* a été séparée en trois espèces distinctes (voir les références dans Costa et al., 2014).

Les Triatominae présentent une grande diversité d'hôtes dont plus de 1150 espèces de mammifères (Patterson, 1994) réparties en 25 familles principalement présentes dans le bassin amazonien (Teixeira et al., 2006 ; Quartier, 2011) mais également des reptiles et des oiseaux. Cependant, seules les espèces responsables du maintien du cycle du parasite à long terme dans un environnement donné sont considérées comme réservoirs (Noireau et al., 2009). Ainsi, les oiseaux et les reptiles sont généralement considérés comme réfractaires à l'infection par *T. cruzi*. Pourtant, il a été observé un cycle complet de *T. cruzi* dans des lézards s'étant nourris à partir de Triatominae infectés (Teixeira et al., 2006).

Trois types d'environnement des vecteurs de la maladie de Chagas peuvent être distingués : un environnement sylvatique, un environnement péri-domicilié et un environnement domicilié. Les vecteurs dits « sylvatiques » vivent dans leur milieu sauvage d'origine et se nourrissent principalement sur des vertébrés non humains (principalement des oiseaux, des mammifères et plus rarement des amphibiens et des reptiles). Ils sont associés à des niches écologiques diversifiées tels que des nids ou des terriers d'hôtes, des creux d'arbres ou de palmiers, des pierriers ou encore des broméliacées ou des cactées (Emperaire and Romana, 2006; Gaunt and Miles, 2000; Sarquis et al., 2011). Généralement les niches

écologiques sont partagées avec leurs hôtes vertébrés. Les espèces du genre *Rhodnius* sont majoritairement associées aux palmiers, tandis que celles du genre *Panstrongylus* le sont le plus souvent aux terriers et cavités d'arbres, et celles du genre *Triatoma*, aux habitats pierreux et aux terriers (Dujardin et al., 1999; Lazzari et al., 2013). L'environnement péri-domicilié correspond aux bâtiments construits pour les animaux domestiques. Les vecteurs domiciliés sont capables d'établir leurs colonies au sein des habitations humaines. La connexion entre ces trois milieux peut se faire par des animaux synanthropiques (opossums, chauves-souris, souris, rats...) qui migrent de plus en plus des environnements sylvatiques vers les environnements humains du fait de la destruction de leurs habitats. Les hôtes vertébrés ainsi que le mouvement des populations humaines peuvent amener les Triatominae à se disperser passivement (Schofield et al., 1999). Mais ces punaises sont également capables de se disperser activement, et ceci principalement par le vol. C'est par ce mécanisme que les Triatominae adultes, ailés, peuvent se déplacer sur de longues distances. Les punaises sylvatiques ou péri-domiciliées peuvent également être attirées par les lumières artificielles de maisons et s'y installer (Coura et al., 2002 ; Lazzari et al., 2013). Ainsi, en présence d'une source lumineuse, *T. infestans* et *R. prolixus* ont une tendance à initier leur vol en direction de cette source, plutôt que vers des points non éclairés (Minoli and Lazzari, (2006); Lazzari et al., 2013). Cette attraction par la lumière joue sans aucun doute un rôle dans les comportements de domiciliation des Triatominae sylvatiques et, par conséquent, a un effet important sur l'épidémiologie de la maladie de Chagas (Barghini and de Medeiros, 2010; Miles et al., 2003; Roux et al., 2011; Walter et al., 2005). La dispersion active par le vol dépend également de la saison, des conditions climatiques, de l'heure, des conditions physiques des individus et de la densité des populations (Lazzari et al., 2013). Ainsi, *T. infestans* semble avoir une activité de dispersion plus importante dans l'heure qui suit le coucher du soleil, particulièrement si les insectes sont à jeun (Vazquez-Prokopec et al., 2004). Des résultats similaires ont été trouvés pour *T. sordida* et *T. guasayana* (Noireau and Dujardin, 2001). Une autre étude a démontré que la dispersion par le vol chez *T. infestans* avait une probabilité plus élevée lorsque la température dépassait les 24°C (Gurevitz et al., 2006). L'âge des adultes ainsi que la disponibilité des hôtes sur le lieu de départ auraient également des effets importants sur l'initiation d'un vol chez cette espèce. Cette étude a aussi mis en évidence que la dispersion par le vol chez *T. infestans* était plus élevée chez les femelles que chez les mâles. Par contre, l'étude de Minoli and Lazzari, (2006) ne relève pas de différence significative de dispersion par le vol entre les femelles et les mâles chez *R. prolixus*. Des résultats similaires ont été trouvés pour les espèces *T. pseudomaculata* et *T. brasiliensis* (Carbajal de la Fuente et al., 2007) qui se dispersent par le vol en proportions équivalentes pour les mâles et les femelles. Par ailleurs, le vol n'est pas le seul moyen de dispersion actif des Triatominae. Ainsi, les femelles de *T. infestans* sont également capables de se disperser en marchant (Abraham et al., 2011). Ce mode de déplacement pourrait être une alternative au vol pour les femelles qui portent leurs œufs et des réserves de sang.

Malgré la dispersion relativement importante des punaises, la plupart des études montrent un isolement par la distance entre les différentes populations d'une même espèce (Gourbière et al., 2012). A courtes distances, les conclusions diffèrent en fonction de l'espèce, de la localité et de l'historique du traitement de pesticides (Gourbière et al., 2012). Ainsi, les populations de *T. infestans* sont généralement très structurées (Breniere et al., 1998; Dujardin et al., 1987; Gourbière et al., 2012), ce qui était attendu pour une espèce essentiellement domiciliée. Cependant, dans les régions où *T. infestans* possède des populations sylvatiques, les flux de gènes entre ces populations et les populations domiciliées sont également restreints voir inexistants (Richer et al., 2007). En revanche, les populations de *T. dimidiata* montrent des flux de gènes importants entre les maisons d'un même village ainsi qu'entre les différents biotopes qui entourent un village (sylvatique, péri-domicilié et domicilié ; Dumonteil et al., 2007). L'analyse de la structure génétique de *T. brasiliensis* basée sur le gène *cytochrome b* a révélé une structuration génétique à la fois à une échelle écologique et à une échelle géographique, la structuration étant moins marquée entre les populations domiciliées (Almeida et al., 2008). D'après cette même étude, la re-infestation des habitations par *T. brasiliensis* après traitement d'insecticide proviendrait de plusieurs populations distinctes. Par ailleurs, une étude récente basée sur l'utilisation de marqueurs microsatellites met également en évidence, à une échelle microgéographique, un brassage génétique entre populations d'origines différentes notamment sylvatiques et domestiques (Almeida et al., submitted).

Trois genres *Triatoma*, *Panstrongylus* et *Rhodnius* et environ 20 espèces de Triatominae sont d'importance épidémiologique (Gourbière et al., 2012). Parmi elles, les espèces domiciliées, ont un rôle majeur dans la transmission de la maladie à l'homme. En effet, la domiciliation de ces espèces implique l'établissement de colonies dans les maisons et le développement du cycle complet de l'insecte en milieu domestique. Cette adaptation au milieu anthropique augmente considérablement la fréquence des contacts des vecteurs avec l'homme. Ainsi, les espèces *R. prolixus*, *T. dimidiata*, *T. infestans*, *T. brasiliensis* et *Panstrongylus megistus* sont les principaux vecteurs de la maladie de Chagas (Figure 3 ; Gourbière et al., 2012).

Figure 3 : Répartition des espèces de Triatominae de grande importance épidémiologique. Seules 20 espèces de Triatominae environ sont responsables de la transmission de *T. cruzi* à l'homme, en raison de leur capacité à infester et, pour certaines espèces, à coloniser l'habitat humain. Les zones rouges indiquent la répartition géographique approximative des espèces. Les espèces écrites en rouge sont considérées comme les plus importants vecteurs du parasite. D'après Gourbière et al., 2012 (cf Figure 1 Gourbière, S., Dorn, P., Tripet, F., and Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* 108, 190–202. : <http://www.nature.com/gate1.inist.fr/hdy/journal/v108/n3/full/hdy201171a.html>)

5. Cas des espèces du genre *Rhodnius* et du complexe d'espèces *T. brasiliensis*

Les *Rhodnius* sont les Triatominae les plus représentés en Amazonie (Abad-Franch and Monteiro, 2007). Les espèces du genre *Rhodnius* peuvent présenter des morphologies similaires (Figure 5). Elles se distinguent des autres Triatominae par l'insertion apicale des antennes et la présence de callosité derrière les yeux qui leur est spécifique (Lent and Wygodzinsky, 1979). Le genre est divisé en quatre complexes *prolixus*, *pictipes*, *pallenscens* et *dalessandroi* (Abad-Franch et al., 2010; Abad-Franch and Monteiro, 2007; Schofield and Dujardin, 1999). Le complexe *prolixus* regroupait initialement les quatre espèces *R. prolixus*, *R. robustus*, *R. nasutus* et *R. neglectus*. A ce complexe, les deux espèces nouvellement décrites *R. montenegrensis* (Rosa et al., 2012) et *R. baretti* (Abad-Franch et al., 2013) proches de *R. robustus* sont à inclure. Le complexe *pictipes* contient quant à lui les espèces *R. pictipes*, *R. amazonicus*, *R. zeledoni*, *R. parensis*, et *R. stali*. Les espèces *R. pallenscens*, *R. colombiensis* et *R. ecuadoriensis* composent le complexe *pallenscens*. Enfin *R. milesi* et *R. dalessandroi* se regroupent dans le complexe *dalessandroi*. Les espèces *R. domesticus*, *R. brethesi*, et *R. neivai* ne sont directement affiliées à aucun des complexes cités ci-dessus. Les espèces présentent des répartitions différentes ; ainsi, à l'est des Andes on retrouve les espèces *R. pictipes*, *R. stali*, *R. brethesi*, *R. amazonicus* et *R. paraensis* tandis qu'à l'ouest on distingue les espèces *R. pallenscens*, *R. ecuadoriensis* et *R. colombiensis*. Le groupe « *robustus* » est principalement retrouvé à l'ouest des Andes et est composé de quatre espèces distinctes : *R. prolixus*, *R. robustus*, *R. neglectus*, et *R. nasutus* (Abad-Franch et al., 2009). Les espèces du genre *Rhodnius* sont principalement arboricoles et généralement rencontrées dans la couronne des palmiers. Mais certaines d'entre elles sont parvenues à coloniser l'habitat humain (*R. prolixus*, *R. neglectus*, *R. nasutus*, *R. stali*, *R. pallenscens* et *R. ecuadoriensis* ; Quartier, 2011). Les foyers de domiciliation de ces espèces, notamment ceux de *R. stali*, se situent dans les régions arides où les moyennes pluviométriques sont très basses (Sombroek, 2001). Ces observations suggèrent que les populations de *Rhodnius* capables de domiciliation dériveraient de populations ancestrales adaptées à des climats arides bien avant l'arrivée des premiers humains sur le continent américain (Abad-Franch and Monteiro, 2007). En suivant cette hypothèse, les populations de vecteurs adaptées à des environnements humides et stables (comme les couronnes de palmiers de forêt tropicale) auraient une faible tolérance aux conditions plus sèches des microclimats des habitations humaines. Par conséquent la probabilité de domiciliation pour ces populations serait plus faible (Quartier, 2011). Cependant la construction des toits en palmes en région tropicale pourrait favoriser le transport passif des Triatominae sylvatiques vers les habitations humaines (Gamboa, 1973). *R. prolixus*, très largement domiciliée, est le vecteur principal de la maladie de Chagas en Colombie, au Venezuela et dans certains pays d'Amérique Centrale. Seules quelques populations sylvatiques de cette espèce ont pu être répertoriées au Venezuela (Felicangeli et al., 2003; Fitzpatrick et al., 2008; Sanchez-Martin et al., 2006) ou en Colombie (Guhl et al., 2009; Rendón et al., 2015)

mais des problèmes d'identification des espèces ne sont pas à écarter . En effet, *R. prolixus* présente des caractéristiques morphologiques très proches de *R. robustus* rendant leur distinction très difficile (Quartier, 2011), et ces deux espèces ne présentent pas de marqueurs alloenzymatiques diagnostiques (Harry, 1993a; Harry et al., 1992) ou très peu de différences génétiques au niveau de marqueurs nucléaires (Monteiro et al., 2000). L'utilisation de séquences mitochondriales (*cytochrome b*) a cependant permis de différencier ces deux espèces et d'identifier quatre clades géographiquement séparés au sein de l'espèce *R. robustus* (Monteiro et al., 2000). Pavan et al., (2013) ont également mis en évidence parmi sept amplicons testés, un SNP (single-nucleotide polymorphism) permettant de distinguer les deux espèces *R. robustus* et *R. prolixus*.

Figure 5 : Espèces appartenant au genre *Rhodnius*

1 à 4 : *R. nasutus* (Dias et al., 2014); 5 : *R. barreti*; 6 : *R. robustus* (Abad-Franch et al., 2013); 7 : *R. pallescens* (<http://www.boldsystems.org>); 8 : *R. montenegrensis* (Meneguetti et al., 2015); 9 : *R. colombiensis*; 10 : *R. prolixus*; 11 : *R. brethesi*; 12 : *R. neglectus*; 13 : *R. pictipes* (J.M. Ayala L.); 14 : *R. amazonicus* (Béranger and Pluot-Sigwalt, 2002); 15 : *R. neiva* (J.M. Ayala L.); 16 : *R. stali* (S. Depickère, INLASA, Bolivia)

Le complexe d'espèces *Triatoma brasiliensis* est un groupe monophylétique comprenant quatre espèces, l'une d'elle incluant deux sous-espèces (Costa et al., 2013; Monteiro et al., 2004). On distingue donc *T. brasiliensis brasiliensis*, *T. b. macromelasoma*, *T. juazeirensis*, *T. melanica* et *T. sherlocki* (Figure 6). Ces espèces peuvent être identifiées par leurs caractéristiques morphologiques qui ont donné lieu à une clé de taxonomie (Costa et al., 2013).

Figure 6 : Les cinq membres du complexe d'espèce *Triatoma brasiliensis*.

1 : *T. b. macromelasoma* mâle; 2 : *T. b. brasiliensis* femelle, 3 : *T. juazeirensis* mâle ;
4 : *T. melanica* mâle; 5: *T. sherlocki* femelle. Barre = 5 mm. (Costa et al., 2013)

(cf Figs 1-5 Costa, J., Correia, N.C., Neiva, V.L., Gonçalves, T.C.M., and Felix, M. (2013). Revalidation and redescription of *Triatoma brasiliensis macromelasoma* Galvão, 1956 and an identification key for the *Triatoma brasiliensis complex* (Hemiptera: Reduviidae: Triatominae). Mem. Inst. Oswaldo Cruz 108, 785–789 : http://www.scielo.br/scielo.php?pid=S0074-02762013000600785&script=sci_arttext&tlng=es)

Le complexe se répartit dans 12 états brésiliens dans les régions de Caatinga et de Cerrado (Costa et al., 2003). Les espèces du complexe diffèrent par leur importance épidémiologique, leur écologie, leur capacité à se disperser et par leurs caractéristiques génétiques (Costa et al., 2014, 2009, 1997).

T. b. brasiliensis, nommée dans la suite du texte *T. brasiliensis* et ayant fait l'objet de notre étude, est le membre le plus étudié du complexe en termes d'épidémiologie et de distribution spatiale : ce vecteur est présent dans six états, utilise diverses sources alimentaires, et est réparti sur plusieurs écotopes. Malgré les réductions significatives de l'incidence de la maladie de Chagas au Brésil au cours des dernières décennies, *T. brasiliensis* reste une préoccupation de santé publique majeure (Costa et al., 2014; Silveira and Vinhaes, 1999). En

effet, dans les environnements sylvatiques, cette espèce vit dans des fissures et des crevasses rocheuses, généralement à proximité de petits mammifères (Sarquis et al., 2010), mais ces Triatominae sont aussi capables de coloniser des milieux anthropiques. Ainsi, ils s'établissent dans les fissures des murs des maisons ou ceux des structures péri-domestiques telles que les enclos, les étables de porcs, les poulaillers, et dans des tas de briques, de tuiles et de bois (Sarquis et al., 2010, 2006). Cette capacité à envahir et à s'adapter à des environnements différents démontre l'importance épidémiologique de cette espèce. Or, en plus du fort taux d'infestation de cette espèce dans les habitations, elle présente un taux d'infection par *T. cruzi* élevé (Costa et al., 1998). *T. brasiliensis* est donc désormais considérée comme le vecteur de la maladie de Chagas le plus important dans les zones semi-arides du nord du Brésil (Costa et al., 2003; Silveira and Vinhaes, 1999). Par ailleurs, l'espèce *T. sherlocki* qui a été décrite pour la première fois en 2002 et reconnue comme un cinquième membre du complexe *T. brasiliensis* (Costa et al., 2013; Mendonça et al., 2009) a récemment colonisé des domiciles humains dans la région minière de Bahia (Almeida et al., 2009). Ceci met en évidence les changements rapides qui peuvent survenir dans l'écologie de ce complexe d'espèces face à des changements environnementaux drastiques.

Dans ce contexte, comprendre comment les espèces de Triatominae font face à ces changements et quels sont les mécanismes qui entrent en jeu dans leur adaptation à ces nouveaux environnements devient crucial pour l'élaboration de nouvelles méthodes de lutte anti-vectorielle.

II. La maladie de Chagas dans le contexte des changements globaux

1. *Risque d'émergence*

Le risque d'émergence ou de ré-émergence d'une nouvelle maladie vectorielle dépend de trois facteurs : i) l'effet de l'environnement sur les agents pathogènes, les vecteurs et les hôtes ; ii) la diversité et l'abondance des agents pathogènes, des vecteurs et des hôtes dans une région, et iii) la fréquence de contact entre l'homme ou des animaux domestiques avec les réservoirs naturels. Parmi les changements globaux, le réchauffement climatique peut avoir un rôle très important dans l'expansion ou la régression d'un pathogène. En effet, la distribution géographique, la durée du développement, la longévité et la période d'activité des agents pathogènes dépendent fortement de la température (Duvallet, 2006; Githeko et al., 2000; Parham et al., 2015; Thomson, 2014). De plus, avec l'augmentation de la température, la migration de certaines espèces arthropodes vectrices d'agents zoonotiques vers des pays auparavant trop froids devient possible. Les changements climatiques engendrent également des répercussions sur le fonctionnement des écosystèmes, la composition des communautés d'espèces et les cycles de nutrition. Ces pressions de sélection sur les vertébrés réservoirs ainsi que sur les arthropodes vecteurs vont conduire à des modifications dans leur répartition, leur abondance, leur dynamique et leur structure génétique. Les populations d'agents infectieux les mieux adaptées aux nouvelles conditions environnementales ainsi qu'à la nouvelle distribution de vecteurs et d'hôtes vont être sélectionnées, entraînant indirectement une modification de leur virulence (Canini, 2010).

La biologie et l'écologie des vecteurs de la maladie de Chagas sont comme de nombreuses espèces directement liées aux précipitations et aux températures (Guarneri et al., 2003; Luz et al., 1999; Okasha, 1964). Par conséquent, les changements climatiques pourraient avoir des répercussions sur l'aire de répartition géographique des insectes vecteurs. Le modèle de niche écologique prédit un effet hétérogène du changement climatique sur l'aire de distribution des vecteurs de la maladie de Chagas en fonction des espèces. En effet, la répartition géographique de *R. prolixus* pourrait être modifiée et s'étendre notamment en Bolívar (province de Buenos Aires en Argentine), dans l'état d'Amazonas (au nord-ouest du Brésil) et dans l'Etat d'Apure (sud ouest du Venezuela) (Medone et al., 2015), tandis que celle de *T. infestans* (Medone et al., 2015) et celle de *T. brasiliensis* (Costa et al., 2014) resteraient à priori relativement stables.

Par ailleurs, les activités humaines directes ou indirectes à l'origine des changements climatiques ont aussi une incidence sur la baisse globale de la biodiversité (Keesing et al., 2010; Mace et al., 2005). L'érosion de la biodiversité impacte indirectement l'émergence des zoonoses. En effet, plusieurs études s'accordent à dire que les communautés à faible richesse en espèces hôtes réservoirs ont tendance à présenter des niveaux plus élevés de prévalence de zoonose que les communautés à plus grand nombre d'espèces hôtes réservoirs. Ce phénomène est appelé

« l'effet de dilution » (Schmidt and Ostfeld, 2001). L'effet de dilution représente l'importance relative des contacts avec des hôtes non compétents par rapport aux hôtes compétents. L'appauvrissement de la biodiversité peut donc favoriser la transmission des maladies animales à l'homme (Keesing et al., 2010). Peu d'études ont été faites sur le lien entre biodiversité des hôtes et prévalence de la maladie de Chagas. Cependant, les conclusions de (Quartier, 2011) et Levy et al., (2015) vont dans le même sens que les observations citées ci-dessus. Une baisse de diversité des hôtes en milieu domestique ou péri-domestique peut favoriser l'augmentation de la prévalence du parasite responsable de la maladie.

La modification des écotopes naturels des punaises par des actions anthropiques peut également intervenir dans l'épidémiologie de la maladie de Chagas en modifiant la dynamique et la structure génétique des populations de vecteurs par interférence avec les populations péri-domestiques et/ou domestiques (voir paragraphe lutte anti-vectorielle ci-dessous). En effet, face à l'anthropisation de leur habitat, les espèces répondent en jouant sur la plasticité phénotypique (capacité d'un génotype à exprimer différents phénotypes dans différents environnements), en migrant et /ou par sélection de mutations conférant un avantage évolutif ; en d'autres termes, en s'adaptant. Dans le bassin amazonien vingt-sept espèces de triatomines ont été décrites essentiellement sylvatiques (Abad-Franch and Monteiro, 2007; Aguilar et al., 2007). La principale cause de la modification des habitats naturels en Amazonie, est la déforestation liée au développement de l'agriculture. En milieu déforesté, les palmiers du genre *Attalea* prolifèrent rapidement et envahissent les milieux ouverts devenant même un fléau pour les pâturages et les cultures. Dans certaines régions du Brésil, les palmiers du genre *Attalea*, particulièrement *Attalea speciosa* (Babaçu), représentent une ressource extractiviste importante et leur abattage est même interdit par le gouvernement (Quartier, 2011). Ces pratiques de dégradation forestière provoquent la prolifération des habitats de prédilection des *Rhodnius spp* et peuvent donc influencer sur les dynamiques des populations de vecteurs.

Longtemps considérée comme non endémique par les organisations mondiales (OMS) et panaméricaines de la santé publique (PAHO), l'Amazonie a assisté depuis les années 1960 à l'apparition de nombreux cas autochtones de trypanosomiase américaine (Aguilar et al., 2007; Coura et al., 2002; Miles et al., 2003; Teixeira et al., 2001). En 2005, l'Office panaméricain de la santé publique a dressé un bilan de la situation en réunissant tous les acteurs concernés (AMCHA, 2005). Effectivement, des cas d'infection humaine ainsi que des hôtes réservoirs ont été décrits dans tous les pays de la région amazonienne (Colombie, Pérou, Venezuela, Guyane française, Brésil). De plus, le mode de transmission par voie orale est en constante augmentation dans cette région (Coura et al., 2002; Monteiro et al., 2010; Valente et al., 1999). Ce mode de transmission a lieu lorsque les insectes interfèrent avec la chaîne alimentaire lors de la production artisanale de jus de fruits, açaí (*Euterpe olacea*), jus de palme, ou même d'eau, par contamination de ces denrées par des insectes migrants des milieux sylvatiques vers les lieux de production ou de stockage des fruits de palmiers. Les premiers cas de transmission orale ont été décrits au Brésil en 1969 (Shaw et al., 1969) et depuis le succès des campagnes

d'élimination des populations de *Triatoma infestans* et le contrôle des banques de sang, la contamination par voie orale est devenue le mode de transmission le plus courant dans ce pays (Coura et al., 2002; Toso et al., 2011). La Guyane a longtemps été considérée comme une zone à faible risque de maladie de Chagas. Cependant depuis notamment les travaux de Aznar et al., (2004) des mesures ont été prises par les autorités en ce qui concerne la collecte du sang qui a été suspendue, montrant une évolution de la transmission de *T. cruzi* dans ce pays.

D'une façon plus générale, la destruction des habitats naturels des vecteurs augmente les risques de contact homme-vecteur. Si au 20ème siècle il s'agissait de lutter contre une zoonose installée dans les habitations humaines, aujourd'hui, il s'agit d'identifier, de caractériser, et de gérer le risque de transmission d'une enzootie qui circule dans un environnement sans cesse en modification (Romaña et al., 2003). Enfin, la capacité vectorielle, c'est-à-dire l'aptitude d'un vecteur à transmettre un agent infectieux en fonction de sa bio-écologie et des conditions environnementales, est dépendante de la densité des vecteurs. En effet, plus la densité est importante, plus il y a de compétition entre les punaises pour s'alimenter sur le même hôte. Le repas sanguin est alors souvent interrompu, ne provoquant pas la déjection de l'insecte et par conséquent le dépôt du parasite sur l'hôte (Dujardin et al., 2000). Cette relation densité de vecteurs-capacité vectorielle a une conséquence très importante dans le choix des stratégies de lutte anti-vectorielle. Des techniques permettant une simple baisse de la densité des vecteurs pourraient avoir pour conséquence d'augmenter le pouvoir vectoriel des insectes survivants. Il est donc nécessaire d'utiliser des méthodes radicales d'élimination totale des vecteurs. Jusqu'à présent, seul le traitement insecticide atteint cet objectif (Dujardin et al., 2000).

2. La lutte antivectorielle et le processus de domiciliation

Les luttes anti-vectorielles ont commencé dans les années 1940 dans certains pays d'Amérique Latine, mais les programmes nationaux d'envergure n'ont été mis place seulement qu'après les années 1970. Le dépistage à grande échelle des donneurs de sang infectés en Amérique latine a commencé dans les années 1980 après l'apparition du sida (Dias et al., 2002). Ces programmes ont permis la réduction des cas aigus et des cas de nouvelles infections dans les classes d'âge de jeunes, ainsi que des réductions progressives des taux de mortalité dans les zones contrôlées (Dias et al., 2002). L'approche internationale, coordonnée par l'OPS (Organisation Panaméricaine de la Santé), a mené à l'élimination de la transmission de la maladie de Chagas en Uruguay et au Chili, et à sa forte diminution au Brésil et en Argentine (Dias et al., 2002) du fait de l'élimination dans ces pays des principaux vecteurs domiciliés, *R. prolixus* et *T. infestans*. Dans ces pays, du fait de l'absence de populations sylvatiques de ces vecteurs, il semble peu probable que les populations de vecteurs puissent se reconstituer après l'élimination des populations domiciliées (Dias et al., 2002).

Cependant, les espèces sylvatiques ne sont pas à négliger. En effet, les modifications environnementales conduisent à un déplacement des populations d'espèces sylvatiques vers les habitations humaines (Abad-Franch et al., 2009; Abad-Franch and Monteiro, 2007; Romaña et al., 1999). Par exemple, la domiciliation de *Triatoma infestans* et de *Rhodnius prolixus* vient

probablement du fait que durant les deux ou trois derniers siècles les pratiques d'agriculture extensive en Amérique Latine ont déplacé les populations d'hôtes incitant les Triatominae à trouver leurs ressources alimentaires dans les habitations humaines (Coura, 2007). Avec les pressions fortes liées aux modifications environnementales actuelles, notamment la déforestation, d'autres espèces de Triatominae d'origine sylvatique colonisent la niche écologique domestique libérée par l'élimination chimique des populations de *T. infestans* ou de *R. prolixus* (Costa et al., 2003; Fitzpatrick et al., 2008; Guhl et al., 2005; Noireau et al., 2000). Par ailleurs, lorsque les populations de Triatominae sont présentes à la fois dans des écotopes domestiques, péri-domestiques et sylvatiques, les flux de gènes entre écotopes sont importants. Les colonies peuvent alors être reconstruites rapidement par des migrants, ce qui limite l'efficacité des pulvérisations d'insecticides (Barbu et al., 2009; Dumonteil et al., 2004; Gourbière et al., 2012). C'est pourquoi les espèces sylvatiques représentent un nouveau défi pour la lutte anti-vectorielle (Costa and Lorenzo, 2009; Gourbière et al., 2012; Guhl et al., 2009) et qu'il est nécessaire d'étudier leur écologie et leur comportement afin de comprendre le fonctionnement du processus de domiciliation et d'en tenir compte dans l'élaboration de nouvelles stratégies de lutte anti-vectorielle (Guhl et al., 2009).

En raison des différences de capacité vectorielle entre les espèces hématophages, leur identification taxonomique est une première étape importante pour la mise en place des programmes de surveillance et de contrôle vectoriel. Les caractères morphologiques et les différents outils cités précédemment lors de la description des vecteurs sont alors utilisés.

L'étude de la diversité génétique au sein des espèces est également essentielle à la compréhension de l'épidémiologie de la maladie de Chagas et par conséquent, à la conception des méthodes de contrôle. Les études de flux géniques permettent d'établir les sources de ré-infestation possibles après le traitement insecticide et peuvent ainsi aider à définir la zonegéographique à traiter lors des campagnes chimiques de lutte antivectorielle afin d'augmenter leur efficacité. La génétique des populations, par le biais d'allozymes, d'ADN mitochondrial et nucléaire, a permis de mettre en évidence que *T. infestans* et *R. prolixus* sont deux espèces domiciliées qui présentent globalement une faible diversité génétique au sein de leurs populations naturelles ainsi que dans les colonies de laboratoire (Dujardin et al., 1998; Harry et al., 1992; Lopez et al., 1995; Pereira et al., 1996). La réduction de la diversité génétique, en raison d'effets fondateurs lors de la domiciliation, pourrait expliquer la faible diversité génétique de leurs populations (Gourbière et al., 2012; Harry et al., 1993a). En revanche, *T. dimidiata* et *T. brasiliensis* sont des espèces répandues dans les habitats domestiques, péri-domestiques et sylvatiques et dont les populations présentent généralement une plus grande diversité. Particulièrement, la ré-infestation des habitats domestiques après traitement chimique par des individus issus des populations sylvatiques diversifiées favorise le brassage génétique et peut également permettre la propagation des gènes de résistance (Barbu et al., 2009; Dumonteil et al., 2004; Gourbière et al., 2012). De tels phénomènes pourraient donc menacer les efforts de lutte vectorielle.

III. L'apport du séquençage à haut débit en biologie de l'évolution

Dans les années 2000, l'apparition des nouvelles générations de séquençage (NGS) permettent de traiter en parallèle des dizaines de milliers de séquences. Cette parallélisation permet de produire un énorme volume de données à moindre coût. Les différentes techniques de séquençage à haut débit (HTS pour « High throughput sequencing ») sont exposées en Annexe. Ces technologies offrent de nouvelles perspectives d'étude en écologie, en génétique et en biologie de l'évolution (Ekblom and Galindo, 2011). Leur utilisation pourrait donc ouvrir de nouvelles pistes dans l'étude du processus adaptatif des vecteurs de la maladie de Chagas aux anthroposystemes.

1. Séquençage complet des génomes

Les coûts du séquençage diminuant de jour en jour, et les technologies s'améliorant rapidement, différentes espèces modèles, ainsi que de nombreuses espèces d'intérêt économique, médicale ou écologique, ont pu être séquencées. Le séquençage complet des génomes a permis d'obtenir des contigs se rapprochant du chromosome complet et permettant l'élaboration de cartes génomiques complètes, ainsi que la révélation du contexte physique des gènes (Werner, 2010). En septembre 2015, on comptait 13468 génomes séquencés et accessibles sur NCBI. Ces génomes proviennent aussi bien d'eucaryotes que de procaryotes, de virus, mais aussi de plasmides ou d'organelles. On dénombre 183 génomes d'insectes.

L'apport des HTS a fourni des informations cruciales sur les variations nucléotidiques. Les HTS ont également permis la découverte à grande échelle de marqueurs génétiques permettant l'amélioration de l'étude des bases génétiques de la variation phénotypique grâce à la cartographie ou des études d'association à l'échelle du génome (GWAS). Les études portant sur l'histoire des populations, leur structure génétique ou l'inférence de parenté ont également été grandement améliorées grâce aux données fournies par cette technologie. Le scan de génome entier et la génomique comparative utilisant des données HTS fournissent de meilleures chances d'identifier des *loci* sous sélection.

Ainsi, de nombreuses études sur l'adaptation des espèces modèles en écologie ont été réalisées par le biais de ces méthodes. Par exemple, les HTS ont permis de déterminer les gènes responsables de l'adaptation d'*Arabidopsis lyrata* aux sols riches en métaux ; de comprendre l'origine des différences morphologiques et écologiques des corégones (poissons des lacs) en fonction de leur habitat ou encore d'évaluer les gènes impliqués dans les variations géographiques et fonctionnelles du pattern des ailes du papillon *Helixionius melpomene* (Stapley et al., 2010). Par ailleurs, les HTS ont également permis l'essor des études écologiques sur les espèces non modèles. En effet, avec le nombre croissant d'espèces séquencées, l'étude génomique des espèces non-modèles est facilitée par la possibilité de se baser sur des ressources génomiques d'espèces apparentées (Ekblom and Galindo, 2011). Par ailleurs, le séquençage à haut débit offre également la possibilité de réaliser des études génomiques sur de nombreuses espèces non-

modèles écologiquement intéressantes et ce sans génome d'un organisme proche génétiquement (Stapley et al., 2010).

2. Le cas du RNAseq

Le séquençage à haut débit peut également être utilisé pour du séquençage de transcriptomes, par la retro-transcription de l'ARN en ADN complémentaire (ADNc). Le séquençage à haut débit de données transcriptomiques est nommé RNAseq. L'ARN est d'abord extrait à partir d'une cellule, d'un tissu, d'un individu complet ou même d'un groupe d'individus. Le transcriptome peut être total (dans ce cas, il contient tous les types d'ARN) ou fractionné (par exemple, le fractionnement par queue poly A permettra la sélection des ARNm). L'ARN est ensuite converti en une librairie d'ADN complémentaire par rétro-transcription. Enfin, la librairie est séquencée à haut débit par une des méthodes énoncées ci-dessus. Ces méthodes ne permettent pas de séquencer les ARN en une seule fois du fait de leur longueur ; c'est pourquoi chaque ARN est fragmenté en petites séquences dont une portion de taille identique entre fragments appelée « read » ou « lecture » est séquencée. Ces reads vont ensuite être assemblés entre eux afin de reconstituer au plus près les chaînes d'ARN sur toute leur longueur. Le regroupement de plusieurs reads en une seule et longue séquence s'appelle un contig. L'assemblage est un puzzle basé sur plusieurs millions de reads qui nécessite de gérer des fichiers de taille très conséquente. Cette étape peut donc s'avérer délicate et ne doit pas être négligée. En effet, si l'assemblage est fragmenté ou chimérique, il peut conduire à des biais dans les analyses ultérieures. Il est donc primordial de générer un assemblage le plus proche possible de la réalité. Deux méthodes principales peuvent être utilisées. La première consiste à aligner les lectures sur un génome ou un transcriptome de référence déjà connu. La seconde, consiste à assembler les reads *de novo* en utilisant le chevauchement des reads. La méthode d'assemblage basée sur une référence présente des difficultés, notamment lors de l'étape de l'alignement. Plus le read est petit, plus la probabilité qu'il s'aligne à plusieurs endroits dans le génome est élevée, et ce d'autant plus qu'il y existe des régions répétées dans le génome. Ensuite, les ARNs messagers matures contiennent uniquement des exons, ainsi, certains reads vont donc recouvrir la fin d'un premier exon et le début de l'exon suivant. L'intron situé entre les deux exons ne sera couvert par aucun read (Figure 7). Lors de l'alignement des reads sur un génome de référence, il faudra donc prendre en compte ce phénomène de maturation et accepter des gaps de la taille d'un intron. Enfin, lors de la maturation, des phénomènes d'épissage alternatif peuvent avoir lieu. Ainsi, pour un même gène, plusieurs isoformes peuvent être transcrits (Figure 7). Les méthodes d'assemblage devront donc prendre en compte ces différentes possibilités et les inclure dans le transcriptome final.

Figure 7 : Alignement sur un génome de référence des reads séquencés à partir de deux isoformes issus d'un même gène (bioinfo.net).

(cf <http://bioinfo-fr.net/analyse-des-donnees-de-sequencage-a-arn>)

Cette méthode complète avantageusement les informations déduites du séquençage de génomes : le RNAseq peut en effet révéler l'emplacement exact des limites de transcription d'un gène. Il pourra également donner des informations sur la présence d'épissage alternatif ainsi que sur les possibles variations de séquence (par exemple, les SNP) dans la région du transcrit (Wang et al., 2009). Par exemple, le séquençage du génome humain a révélé la présence de 23 000 gènes quand les évaluations antérieures estimaient 100 000 gènes. Cette valeur proche du nombre de gènes de *C. elegans* (20 000) et de *D. melanogaster* (14 000) a soulevé beaucoup de questions sur le rapport entre la complexité du génome et les différences entre organismes (Werner, 2010). Des études de génomique fonctionnel menées en parallèle du séquençage complet des génomes (Werner, 2010) ont pu corrélérer la complexité fonctionnelle non pas avec le nombre de gènes mais plutôt avec la complexité de la régulation de l'expression de ces gènes (transcriptions alternatives, modifications post-traductionnelles...).

Si l'espèce étudiée ne possède pas de référence ou si celle-ci est fragmentée ou de mauvaise qualité, un assemblage *de novo* peut être effectué (Wang et al., 2009). Ainsi, contrairement aux approches basées sur l'hybridation, le RNAseq ne se limite pas à la détection de transcrits correspondant à la séquence génomique existante. De plus, ne nécessitant pas forcément de génome de référence, le RNAseq est particulièrement attrayant pour les organismes non-modèles avec des séquences génomiques qui sont encore à déterminer. L'assemblage *de novo* consiste à utiliser les chevauchements de séquences entre différents reads. En raison de la possibilité de transcrits sens et antisens et de la variation importante d'expression entre gènes conduisant à un recouvrement non uniforme, l'assemblage *de novo* de données transcriptomiques présente des défis spécifiques non rencontrés dans l'assemblage génomique. Par ailleurs, les isoformes, les différentes formes alléliques ou encore les répétitions dans le génome donnant lieu à plusieurs transcrits partiellement semblables vont engendrer plusieurs chemins de recouvrement possibles. Plus ils seront nombreux, et plus il sera difficile de reconstruire et de démêler toutes les possibilités (Sammeth, 2009). Les assembleurs utilisent généralement l'un des deux algorithmes de base : les graphiques de chevauchement et les graphiques de Bruijn. Les graphiques de chevauchement sont utilisés par la plupart des assembleurs conçus pour du séquençage générant des longs reads. Les chevauchements entre chaque paire de reads sont calculés et compilés dans un graphique, dans lequel chaque nœud représente un read unique. Cet algorithme nécessite plus de calculs que les graphiques de Bruijn. Il est plus efficace lorsque les reads sont peu nombreux mais avec un degré de chevauchement élevé. Pour assembler des séquences plus courtes (environ 100pb) et plus nombreuses, les graphiques de Bruijn alignent des k-mers (oligonucléotides de taille k, k variant généralement entre 25 et 50 pb) sur la base de conservation de séquence de k-1 (Figure 8c). L'utilisation de k-mers dans les graphiques de Bruijn réduit l'importance de calcul de cette méthode.

Avec le développement du RNAseq, de nombreux logiciels d'assemblage se sont développés en faisant appel à des méthodes plus ou moins similaires. Par exemple, parmi les assembleurs *de novo* de reads longs, de nombreux logiciels sont disponibles. Parmi eux, on retrouve des logiciels comme MIRA (Chevreux et al., 2004), CAP3 (Huang and Madan, 1999), Newbler (Knudsen et al., 2007), NGen (<http://www.dnastar.com/t-nextgen-seqman-ngen.aspx>), ou CLCbio (<http://www.clcbio.com>). Pour assembler des reads courts généralement produits par la technologie Illumina, divers logiciels sont disponibles, parmi lesquels certains sont spécifiquement consacrés aux données transcriptomiques : Trans-Abyss (Robertson et al., 2010), Trinity (Figure 8, Grabherr et al., 2011), Oasis (Schulz et al., 2012), Rnnotator (Martin et al., 2010), Multiple-K (Surget-Groba and Montoya-Burgos, 2010) ou SOAPdenovo-Trans (Xie et al., 2013)... Pour effectuer un assemblage basé sur une référence, les deux logiciels les plus utilisés sont : Cufflink (Trapnell et al., 2010) et Scripture (Guttman et al., 2010). Par ailleurs, l'usage de chaque logiciel est modulable grâce à de nombreuses options. Par exemple, lors d'un assemblage *de novo* par graphe de Bruijn, la taille des k-mers peut être choisie. Lors d'un assemblage sur référence, on peut par exemple guider l'assemblage par l'utilisation de gènes annotés dans la référence ou au contraire ne pas en tenir compte. Cependant à l'heure actuelle, il n'y a pas de consensus clairs sur les meilleurs logiciels et les paramètres à utiliser. Le choix du logiciel d'assemblage et des différents paramètres peuvent donc s'avérer complexes et dépendent en partie du jeu de données à assembler. La qualité de l'assemblage peut être évaluée par l'intermédiaire de plusieurs critères. Pour savoir si l'assemblage reste fragmenté, le nombre de contigs et le N50 (longueur des contigs à partir de laquelle la somme des longueurs des contigs de même taille ou plus grands couvre 50 % de l'assemblage) peuvent être utilisés. Pour évaluer la représentativité du transcriptome, on peut se baser sur des banques de données de protéines très conservées et largement exprimées dans l'ensemble des cellules. Si la majorité des transcrits codant ces protéines est retrouvée, alors on peut considérer que le transcriptome est complet.

Figure 8 : Exemple d'assemblage *de novo* : méthode utilisée par le logiciel Trinity (Grabherr et al., 2011).

(cf Figure 1 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-652 :

<http://www.nature.com/gate1.inist.fr/nbt/journal/v29/n7/abs/nbt.1883.html>)

Contrairement aux puces à ADN qui manquent de sensibilité pour les gènes exprimés à de faibles niveaux ou au contraire à des niveaux très élevés, le RNAseq présente une gamme de détection de gènes bien plus étendue (Wang et al., 2009). Il permet aussi de comparer assez précisément le niveau d'expression d'un gène entre plusieurs conditions (par exemple, entre différents tissus, entre individus ou entre populations ou en fonction des conditions environnementales ; Werner, 2010). La synthèse et la maturation des ARNs sont très contrôlées et forment des réseaux d'expression de gènes complexes qui mènent à l'élaboration

de processus biologiques. Ces réseaux doivent être à la fois robustes et plastiques afin de permettre une rapide adaptation à des changements environnementaux ou à des perturbations génétiques. Le nombre de lectures alignées sur un gène ou un transcrit étant proportionnel à l'abondance de son ARN dans la cellule, le RNAseq offre la possibilité d'analyser l'expression de l'ensemble des gènes et ainsi d'approcher les réseaux d'expression de gènes. Cette méthode ouvre donc de nouvelles opportunités pour mieux appréhender les liens qui relient phénotypes et génotypes (Marguerat and Bähler, 2010).

Lors d'une analyse d'expression par le biais du RNAseq, une première étape d'alignement revient à rechercher dans le génome ou dans le transcriptome de référence la position de la séquence similaire à celle du read. La présence d'isoformes dans les jeux de données complexifie cette étape. En effet, différents transcrits alternatifs peuvent être exprimés à partir d'un même gène. Ils comporteront alors des exons différents mais également des régions communes entre eux. Une lecture dont la séquence correspond à cette partie commune pourra être alignée sur plusieurs transcrits. Il devient alors difficile de savoir à quelle isoforme cette lecture appartient induisant des difficultés dans le comptage. Les tests statistiques d'expressions différentielles consistent à savoir si dans la condition A, l'expression d'un transcrit donné est significativement plus importante que dans la condition B. Différents logiciels d'analyses d'expression différentielle sont utilisés. Deux d'entre eux sont particulièrement recommandés : DESeq2 (Love et al., 2014) et EdgeR (Robinson et al., 2010). Le séquençage à haut débit comporte différents biais qui peuvent fausser les comparaisons de données brutes. En effet, les artefacts de PCR, les différences de taille de banque entre échantillons ou le pourcentage de lectures alignées par échantillon peuvent conduire à des tailles d'échantillons (nombre de lectures alignées) très variables et en conséquence fausser les résultats. Avant de comparer l'expression dans les différents échantillons, les données de comptage doivent donc être rendues comparables. Cette étape s'appelle la normalisation. Pour cela, DESeq2 et EdgeR (normalisation TMM) partent de l'hypothèse que la majorité des gènes ne sont pas différentiellement exprimés. Cependant, bien que l'idée de départ soit la même, les deux packages utilisent des méthodes de normalisation différentes. DESeq2 calcule pour chaque gène ou transcrit une moyenne géométrique sur tous les échantillons afin d'évaluer un facteur de taille qui sert ensuite à normaliser les données de comptage des échantillons. EdgeR calcule une moyenne tronquée des lectures alignées (TMM) entre chaque paire d'échantillons après exclusion des gènes ou des transcrits qui comprennent le plus grand ratio de lectures alignées. A partir de ces données normalisées, DESeq2 et EdgeR évaluent la dispersion de l'expression. DESeq2 utilise une fonction lissée de la dispersion en fonction de l'expression des gènes. EdgeR fonctionne en 3 étapes successives : 1) estimation d'un coefficient de dispersion global sur tous les transcrits, 2) estimation de la variation de la dispersion en fonction de l'abondance des transcrits et 3) ajustement de l'estimation du coefficient de dispersion pour chaque transcrit. Une méthode d'estimation robuste de la dispersion est également proposée dans EdgeR qui permet de réduire l'influence des mesures « outliers » en accordant des poids aux différentes

estimations. A partir des données normalisées et de l'estimation de la dispersion, DEseq2 et EdgeR évaluent si les contigs sont significativement différentiellement exprimés en utilisant une binomiale négative. Les résultats de RNAseq sont proches en termes de qualité de ceux acquis par qPCR et montrent des niveaux élevés de reproductibilité qui attestent de leur fiabilité (Wang et al., 2009).

Ces différentes approches offrent donc des possibilités considérables pour les études en écologie, en génétique et en biologie de l'évolution (Ekblom and Galindo, 2011). Il est donc évident que ces nouvelles technologies sont en train de changer le paysage de la recherche dans ces domaines. Cependant, il reste encore de nombreux défis à relever associés à l'analyse et à l'interprétation des données à haut débit. C'est pourquoi, ces technologies ne sont pas nécessairement une solution simple pour identifier les gènes impliqués dans l'adaptation. Néanmoins, de nombreux progrès ont été réalisés pour répondre à cette demande, et les méthodes d'analyses sont améliorées de jour en jour.

3. Chez les *Triatominae*

Le génome de *Rhodnius prolixus* a été séquencé au Genome Sequencing Center (WUGSC) de l'Université de Washington en Septembre 2010 à partir d'une colonie d'élevage originaire de Colombie et a été publié dans GenBank en Décembre 2010: ACPB000000000.2. Il est composé de 58 560 contigs assemblés en 27 872 scaffolds avec une longueur totale de 702 Mb. L'annotation de ce génome est en cours et les premiers résultats devraient être publiés très prochainement. A l'heure actuelle, 17 244 gènes ont été prédits automatiquement sur ce génome.

Par ailleurs, les sialotranscriptomes de *T. brasiliensis* (Santos et al., 2007), de *T. infestans* (Assumpção et al., 2008), de *T. dimidiata* (Kato et al., 2010), de *Diptellogaster maxima* (Assumpção et al., 2011), de *R. prolixus* (Bussacos et al., 2011) ont été séquencés en Expressed Sequence Tag (EST) dans le but de définir des composés pharmacologiques potentiellement intéressants. Le transcriptome du follicule ovarien de *R. prolixus* a également été séquencé par le biais de la même technologie (Medeiros et al., 2011). Seul le transcriptome du tube digestif de *R. prolixus* a été séquencé à haut débit (Ribeiro et al., 2014). Dans cette étude récente, les ARNs des trois segments distincts de l'intestin d'adultes ont été extraits et séquencés en 454. Le transcriptome a été assemblé en 317 104 contigs dont 27 751 possédait une longueur supérieure à 499 nucléotides. Ces transcrits ont aidé à la prédiction du répertoire de gènes du génome de *R. prolixus* et ont été utilisés comme support à l'annotation manuelle (Ribeiro et al., 2014). Plusieurs transcrits correspondant à des fonctions attendues, tels que les enzymes digestives ou les transporteurs, sont apparus en grand nombre. Cette étude a également apporté de nouvelles données qui pourraient aider à comprendre les aspects de la physiologie digestive de cet insecte et son interaction avec le microbiote intestinal et les Trypanosomatidae.

IV. Le système chimiosensoriel chez les insectes

Une des pistes pour comprendre le processus de domiciliation est de s'intéresser au système chimiosensoriel des Triatominae. En effet, les insectes interagissent entre eux et avec leur environnement notamment *via* des molécules qualifiées de semiochimiques. Les protéines chimiosensorielles qui permettent de détecter les stimuli de l'environnement jouent un rôle très important dans l'adaptation des insectes à de nouveaux milieux et pourraient être impliquées dans le processus de domiciliation. Or, le développement du séquençage complet des génomes par le biais des technologies à haut débit ont permis d'accroître de façon considérable nos connaissances du système chimiosensoriel des insectes.

1. Les molécules semiochimiques

Les molécules chimiques utilisées par les insectes pour communiquer entre eux et percevoir leur environnement sont classées en différentes catégories selon leur nature et leurs effets (Mercier et al., 2013). Ainsi, les phéromones permettent une communication intraspécifique (Karlson and Luscher, 1959). Elles peuvent agir sur le comportement (phéromones incitatrices ou de déclenchement) et/ou sur la physiologie (phéromones modificatrices ou d'induction). Les molécules allélochimiques quant à elles permettent une communication interspécifique.

Certaines molécules semiochimiques sont volatiles, c'est-à-dire qu'elles se dispersent dans l'air par diffusion à partir de la zone d'émission. Elles sont impliquées dans la communication à plus ou moins longue distance et permettent en général la reconnaissance d'un congénère ou d'un hôte. Les molécules non volatiles sont utilisées pour communiquer par contact. En général, les composés volatiles sont détectés par le système olfactif tandis que les composés non volatiles le sont par le système gustatif (Glaser, 2013).

2. Mode de fonctionnement du système chimiosensoriel

A. Mécanismes généraux

L'interaction des insectes avec leur environnement *via* des molécules chimiques nécessite la détection périphérique de ces molécules, leur réception et l'intégration de ces signaux dans les systèmes nerveux de l'insecte. Les structures sensorielles assurant l'olfaction et la gustation sont en majorité localisées sur les antennes et les pièces buccales des insectes, mais aussi sur les pattes où le message chimique est détecté au niveau de sensilles (Glaser, 2013).

Les sensilles olfactives (Figure 9) permettent de détecter la présence de molécules chimiques volatiles (Calatayud et al., 2013). A la surface des sensilles olfactives de nombreux pores permettent aux composés chimiques de pénétrer à l'intérieur de celles-ci et d'accéder à la membrane des neurones chimiorécepteurs insérés dans le corps de la sensille (Schoonhoven et al., 2005). Cependant, les molécules volatiles sont généralement insolubles dans l'eau. Or, la lymphe sensillaire est un milieu aqueux. Pour que ces molécules circulent dans la lymphe et atteignent la membrane des neurones chimiorécepteurs, des protéines solubles vont lier les

molécules odorantes de manière spécifique. Deux types de protéines sont impliqués dans ce transport : les Odorant Binding Proteins (OBPs) et les Chemosensory Proteins (CSPs). Une fois arrivées près des neurones chimiorécepteurs, les molécules odorantes sont libérées et réceptionnées par des protéines insérées au niveau de la membrane dendritique de ces neurones.

On trouve deux catégories de récepteurs : les récepteurs ionotropiques (IR) et les récepteurs olfactifs (OR) associés à des corécepteurs (ORco). Lorsque ces chimiorécepteurs captent une molécule olfactive, ils traduisent le signal chimique en signal électrique par dépolarisation de la membrane dendritique (Figure 7) (Calatayud et al., 2013). Les impulsions sont transmises le long de l'axone du neurone chimiosensoriel et convergent vers les glomérules, structures d'intégration des informations olfactives. Chaque neurone récepteur ne capte qu'un ou que quelques composés chimiques particuliers, et un glomérule reçoit l'information d'un seul type de neurone récepteur.

Figure 9 : Mécanisme du système olfactif dans la sensille (cf Figure 1 Suh, E., Bohbot, J. D., & Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Current opinion in insect science*, 6, 86-92. <http://www.sciencedirect.com.gate1.inist.fr/science/article/pii/S2214574514001084?np=y#>)

En ce qui concerne les mécanismes gustatifs, le fonctionnement est globalement le même. A la différence des sensilles olfactives, les sensilles gustatives possèdent un seul pore terminal ou une couronne de pores à leur extrémité (Schoonhoven et al., 2005). Les molécules impliquées dans le goût peuvent être solubles ou insolubles. Les molécules solubles circulent librement dans la lymphe sensillaire alors que les molécules insolubles circulent à l'aide de protéines de transport, les Gustatory Transport Protéins (GTP). Les mécanismes gustatifs restent encore peu connus. Il semblerait qu'une fois à proximité de la membrane des neurones gustatifs, de la même manière que pour l'olfaction, les molécules gustatives soient captées par des récepteurs gustatifs. Ces derniers seraient alors activés et induiraient des potentiels d'action acheminés par les axones des neurones gustatifs vers le système nerveux central (Glaser, 2013).

B. Les gènes chimiosensoriels

Le système chimiosensoriel joue un rôle majeur dans les différents traits de vie de l'insecte, lui permettant de reconnaître de nombreux composés aussi bien impliqués dans la recherche de nourriture que dans la recherche de partenaires sexuels ou encore dans l'évitement de prédateurs. Ces gènes chimiosensoriels appartiennent à des familles multigéniques. Concernant leur évolution, plusieurs études tendent à valider le modèle naissance/mort (Nei and Rooney, 2005). Dans ce modèle, de nouveaux gènes naissent par duplication de gènes en tandem. L'évolution individuelle de ces copies et l'expansion ultérieure et successive dans le génome concourent à la formation d'une famille multigénique (Kahn, 1985). Les processus évolutifs et la sélection naturelle qui agissent sur les deux copies peuvent conduire à plusieurs scénarios. Si la duplication confère un phénotype létal ou une baisse de la valeur sélective, alors elle ne sera pas maintenue dans les générations suivantes. Si les copies sont évolutivement

stables, alors la question de leur fonctionnalité se pose. La duplication d'un gène peut donner lieu à une copie fonctionnelle ou à un pseudogène (forme non-fonctionnelle du gène). Les pseudogènes peuvent jouer un rôle important dans le réarrangement génomique; et ce d'autant plus facilement et fréquemment qu'ils sont nombreux et proches. Ils participent donc à la dynamique du génome et constituent une réserve de diversité génétique. Par ailleurs, si les deux gènes dupliqués sont maintenus sous forme fonctionnelle, ils vont diverger au fur et à mesure de leur évolution aléatoire respective. La divergence entre les deux copies peut avoir lieu dans la région régulatrice, modifiant alors l'expression du gène ou dans la région codante, conférant alors une nouvelle fonction au gène concerné et développant éventuellement des affinités avec de nouveaux ligands. Par conséquent, le nombre de gènes contenus au sein de chaque famille est principalement le résultat d'événements aléatoires de duplication et de perte de gènes (Sánchez-Gracia et al., 2009). On nomme paralogues les gènes issus de duplication génique présents au sein d'une même espèce et orthologues les copies dérivées d'une même copie ancestrale trouvées chez des espèces différentes.

L'étude de l'évolution des gènes chimiosensoriels a réellement débuté avec la comparaison des génomes de 12 espèces de drosophiles (Clark et al., 2007). Cette étude a révélé en moyenne 49 OBPs, 4 CSPs, 63 ORs et 60 GRs par espèce de *Drosophila*, le nombre de gènes de chaque famille variant peu entre les espèces. Chez les 12 espèces de drosophiles étudiées, les gènes d'OBPs se regroupent en clusters sur les chromosomes (Galindo and Smith, 2001; Hekmat-Scafe et al., 2002; Vieira et al., 2007; Vogt et al., 2002). Ainsi, 69% des gènes codant pour une OBP sont disposés en 10 groupes de 2 à 6 gènes (Sánchez-Gracia et al., 2009) alors que les gènes d'ORs et de GRs semblent plus dispersés (Robertson et al., 2003). L'arrangement en clusters est également observé pour les CSPs, les ORs et les GRs chez plusieurs insectes (Bohbot et al., 2007; Engsontia et al., 2008; Robertson and Wanner, 2006). Les duplications géniques, spécifiques aux différentes espèces, expliquent la grande variation du nombre de copies observé au sein de chaque famille de gènes selon les espèces (Sánchez-Gracia et al., 2009). La variabilité des paralogues (nombre et polymorphisme) rend alors difficile les études des relations d'orthologie entre les espèces. Certains gènes sont cependant mieux conservés que d'autres du fait de la contrainte de sélection exercée sur eux. Ainsi, généralement, les gènes codant les CSPs sont plus conservés que ceux codant les OBPs. Chez la drosophile, la divergence observée est plus faible pour les gènes codant les GRs que pour ceux codant les ORs (Gardiner et al., 2008a).

C. Les protéines de liaison

Les OBPs et les CSPs constituent les deux principales familles de protéines de la lymphe sensillaire des arthropodes (Angeli et al., 1999; Pelosi et al., 2006, p. 200). En 1981, Vogt et Riddiford ont mis en évidence chez le lépidoptère *Antheraea polyphemus* une protéine capable de lier une phéromone de cette espèce : ils avaient découvert pour la première fois l'existence des OBPs. Le premier membre de la famille des CSPs (nommé p10) a été découvert

dans un contexte non lié à la chimioréception. En effet, la protéine p10 a été identifiée pour son implication dans la régénération des membres chez le cafard (Nomura et al., 1992; Nomura Kitabayashi et al., 1998). L'importance dans la chimioréception des CSPs a été suggérée du fait de leur abondance dans les organes olfactifs et gustatifs des insectes (Angeli et al., 1999; McKenna et al., 1994; Pelosi et al., 2014, 2006; Pikielny et al., 1994; Vieira and Rozas, 2011; Wanner et al., 2004).

Les OBPs et les CSPs sont des molécules solubles synthétisées par les cellules accessoires des neurones olfactifs et sont sécrétées dans la lymphe sensillaire où elles sont très abondantes (Pelosi et al., 2006). Les OBPs sont de petite taille (environ 150 acides aminés). Elles sont caractérisées par six cystéines conservées reliées entre elles par des ponts disulfures conférant à la protéine une structure tridimensionnelle spécifique à six hélices alpha. Elles possèdent également un site hydrophobe où les molécules olfactives non solubles peuvent être accueillies (Leal et al., 1999; Sandler et al., 2000; Scalon et al., 1999). Les CSPs sont plus petites que les OBPs (100 à 120 acides aminés) et leur structure tridimensionnelle qui résulte de la conservation de 4 cystéines et de 2 ponts disulfure se distingue de celle des OBPs (Campanacci et al., 2001; Liu et al., 2012; Mosbah et al., 2003; Tomaselli et al., 2006).

Les gènes paralogues et orthologues constituant les familles multigéniques des OBPs et des CSPs évoluent rapidement. Les CSPs sont globalement plus conservées que les OBPs et montrent communément entre 40 et 50% d'identité de séquences entre des espèces appartenant à des ordres différents, alors que cette identité peut être seulement de 8% pour les OBPs (Pelosi et al., 2005). En fonction des espèces, le répertoire des OBPs et celui des CSPs peut être plus ou moins diversifié. Par exemple, chez les Paranéoptères, le génome du puceron du pois *Acyrtosiphon pisum* comprend 18 gènes codant pour des OBPs et 13 pour des CSPs (Vieira and Rozas, 2011; Zhou et al., 2010a); celui du poux *Pediculus humanus* seulement 5 gènes d'OBPs et 7 de CSPs (Kirkness et al., 2010). Chez les diptères, le répertoire en OBPs est très diversifié à l'inverse de celui des CSPs. Ainsi, chez *Drosophila melanogaster*, on compte 51 gènes codant pour des OBPs et seulement 4 pour des CSPs. Les lépidoptères présentent des répertoires très diversifiés à la fois en terme d'OBPs et de CSPs : *Bombyx mori* par exemple possède 46 gènes d'OBPs et 24 de CSPs (Gong et al., 2009; Vieira and Rozas, 2011).

Certaines OBPs sont spécifiques des phéromones et constituent une sous-famille appelées les Pheromones Binding Proteins (PBPs). Elles sont particulièrement abondantes dans les antennes des mâles chez les lépidoptères (Vogt and Riddiford, 1981). Une autre sous famille d'OBPs, retrouvée chez plusieurs insectes (dont la drosophile, l'anophèle et le puceron), présente plus de six résidus de cystéine et sont nommées OBPs 'Plus-C' (Pelosi et al., 2006). La liaison et le transport d'une molécule odorante particulière va être effectuée par une OBP qui lui est spécifique. Ainsi, il a été montré chez *D. melanogaster* que l'inhibition de la Pheromone Binding Protein « LUSH » induisait une perte totale de la sensibilité à la phéromone 11-cis vacényl acétate (Xu et al., 2005). Chez le moustique *Culex quinquefasciatus*, l'inactivation d'une OBP par ARN interférant entraîne une forte diminution de la réponse antennaire à un

composé phéromonal spécifique (Pelletier et al., 2010). Par ailleurs, les OBPs seraient également impliquées dans la présélection des molécules odorantes d'un récepteur par liaison spécifique (Leal et al., 2005). Ainsi, chez *B. mori* le récepteur BmOR-1 exprimé seul répond à la fois aux phéromones bombykol et bombikal. Mais, en présence de la Pheromone Binding Protein 1, ce même récepteur ne répond plus qu'au bombykol. La PBP1 a donc présélectionné une molécule olfactive pour assurer la spécificité du récepteur (Große-Wilde et al., 2006). Un autre rôle attribué aux OBPs serait celui de protéger les molécules odorantes de leur dégradation par les enzymes lymphatiques avant l'interaction avec le récepteur (Kaissling, 2009). Il a également été proposé que les OBPs changeaient de conformation en présence du ligand ce qui conférerait à ce dernier la faculté d'activer le récepteur olfactif (Laughlin et al., 2008). Les OBPs jouent donc des rôles multiples en lien très fort avec la chimioréception. Cette relation étroite est moins claire pour les CSPs.

En plus de leur abondance dans la lymphé sensillaire des antennes, les OBPs et les CSPs sont également exprimées dans de nombreux autres structures, notamment les pattes (Hua et al., 2013; Ji et al., 2013; Mameli et al., 1996; Picimbon et al., 2001), les ailes (Ban et al., 2003; Hua et al., 2013; Ji et al., 2013), le cerveau (Ji et al., 2013; R. Liu et al., 2012), l'abdomen (Hua et al., 2013; Ji et al., 2013) et les glandes à phéromones (Dani et al., 2011; Gu et al., 2013; Jacquín-Joly et al., 2001; Li et al., 2008; Sun et al., 2012). Ces protéines de liaisons sont donc probablement impliquées dans d'autres processus que l'olfaction (Pelosi et al., 2014). Ainsi, chez l'abeille *Apis mellifera*, certaines CSPs seraient impliquées dans le développement de l'ectoderme des embryons (Forêt et al., 2007; Maleszka et al., 2007). Récemment, les CSPs ont également été identifiées comme des fixateurs potentiels d'insecticides permettant aux insectes de résister aux effets létaux du poison (Liu et al., 2014; Xuan et al., 2015). Plus récemment, les OBPs ont également été liées à d'autres rôles que la chimioréception (Pelosi et al., 2014) : les résultats de Ishida et al., (2013) suggèrent que l'OBP56a, exprimée dans le disque oral de la mouche *Phormia regina*, solubiliserait les acides gras de la viande fraîche et permettrait leur transport vers l'intestin moyen où l'environnement de faible pH facilite la libération et l'absorption ultérieure des acides gras, indispensable à la reproduction de la mouche. Deux autres exemples d'OBP ayant des fonctions non liées à l'olfaction sont retrouvés chez le moustique *Aedes aegypti*. Chez ce moustique, une OBP est impliquée dans des processus inflammatoires (Calvo et al., 2009) et une autre est retrouvée dans les ovaires et les ovocytes où elle serait impliquée dans les mécanismes de reproduction responsables de la maturation ovocytaire (Costa-da-Silva et al., 2013; Marinotti et al., 2014).

D. Les récepteurs

C'est à la fin du siècle dernier que les premiers Récepteur Olfactifs (OR) ont été découverts par l'intermédiaire d'études bioinformatiques sur le génome nouvellement séquencé de *D. melanogaster* (Clyne et al., 1999; Gao and Chess, 1999; Vosshall et al., 1999). Depuis, ils ont été décrits chez de nombreux autres insectes. Codés également par des gènes appartenant à

des familles multigéniques, les ORs sont très divergents d'une espèce à l'autre, mais aussi au sein de la famille multigénique d'une seule et même espèce (Jacquin-Joly and Merlin, 2004). A ce jour, peu d'études fonctionnelles ont démontré l'association OR-ligand. Néanmoins, les quelques études menées ont permis de montrer que certains ORs possédaient des affinités variables pour les molécules odorantes, certains étant spécifiques d'une molécule particulière, d'autres présentant un spectre de reconnaissance plus large (Hallem et al., 2004; Tanaka et al., 2009). Une même molécule odorante peut activer plusieurs récepteurs, et un même récepteur peut reconnaître plusieurs molécules. Ainsi, le nombre de combinaisons possibles permet de reconnaître une grande variété d'odorants (Glaser, 2013). Parmi les ORs, un groupe particulier est impliqué dans la reconnaissance des composés phéromonaux, les PRs. Ces récepteurs, contrairement à la majorité des autres ORs, sont spécialisés dans la réception d'une seule phéromone et y sont très sensibles (Wilson and Mainen, 2006). Un autre type d'ORs se distingue : les ORco (corecepteurs). Ils sont retrouvés chez toutes les espèces d'insectes étudiées et sont très conservés (Krieger et al., 2003; Malpel et al., 2008). Ils sont exprimés dans la quasi-totalité des neurones olfactifs mais ne confèrent de sensibilité à aucun odorant lorsqu'ils sont exprimés seuls (Elmore et al., 2003). Leur rôle réside en la formation d'hétérodimères avec un autre OR. Cette association est indispensable au second OR pour son activation. Cet hétérodimère forme la véritable unité moléculaire fonctionnelle de la reconnaissance olfactive (Benton et al., 2006; Larsson et al., 2004).

En 2009, Benton et al. ont découvert une autre famille de récepteurs olfactifs : les récepteurs ionotropiques (IRs). Ces protéines présentent 3 acides aminés au niveau du site actif de liaison au ligand, suggérant une interaction avec d'autres composés (Glaser, 2013). Les IRs sont classés en deux sous catégories : les IRs antennaires assez conservés jouant un rôle dans l'olfaction et les IRs gustatifs, moins conservés, impliqués dans la gustation (Croset et al., 2010). Chez la drosophile, certains IRs ont été proposés comme jouant le rôle de corécepteur (Abuin et al., 2011; Croset et al., 2010). Chez la même espèce, un IR activerait le comportement de cour chez les mâles en présence d'odeur de fruit, reliant recherche alimentaire et recherche de partenaire sexuel (Grosjean et al., 2011).

Avec le séquençage du génome de la drosophile, les récepteurs gustatifs (GRs) ont également pu être découverts. Ces protéines exprimées dans les organes gustatifs ont des motifs partagés avec les ORs et sont regroupées dans la même superfamille (Clyne et al., 2000). Cependant, ils montrent une divergence encore plus marquée que celles ORs. Chez la drosophile, les ligands de certains GRs ont été caractérisés. Ainsi, on trouve des récepteurs aux sucres (Chyb et al., 2003; Dahanukar et al., 2001; Slone et al., 2007), aux substances amères (Lee et al., 2009; Moon et al., 2006; Weiss et al., 2011), aux phéromones (Miyamoto and Amrein, 2008) ou aux insecticides synthétisés par des plantes (Lee et al., 2012).

3. Le système chimiosensoriel chez les *Triatominae*

A. La communication chimique

La communication chimique des *Triatominae* intervient dans les phénomènes d'agrégation, d'accouplement et de signal d'alarme (Lazzari et al., 2013). Ainsi, les fèces de *Triatominae* sont une source de phéromones d'agrégation qui attirent les insectes et les incitent à s'agréger à leur proximité (Figueiras et al., 1994; Schofield and Patterson, 1977). Ces signaux chimiques pourraient également être impliqués dans l'indication d'une source d'alimentation appropriée (Schofield and Patterson, 1977) et des chemins d'accès aux refuges suggérant un rôle important de ces phéromones dans la sélection des abris anthropiques (Lorenzo and Lazzari, 1996).

Les données récentes sur l'étude des phéromones sexuelles des *Triatominae* suggèrent qu'elles sont synthétisées dans les glandes metasternales femelles et sont préférentiellement émises la nuit (Pontes et al., 2008) incitant les mâles à quitter leurs abris, à s'envoler (Zacharias et al., 2010) et à s'orienter vers les courants d'air chargés en phéromones (Lazzari et al., 2013; May-Concha et al., 2013; Vitta and Lorenzo, 2009).

Les *Triatominae* libèrent une odeur âcre lorsqu'elles sont dérangées. Ces sécrétions produites par les glandes de Brindley jouent le rôle de phéromone d'alarme quand elles sont en forte concentration et repoussent les punaises (Manrique et al., 2006). Cependant, à faible concentration, les punaises s'approchent de la source (Guerenstein and Guerin, 2001; Ward, 1981). De plus, la glande de Brindley n'est fonctionnelle qu'au stade adulte (Lazzari et al., 2013). Par conséquent, les juvéniles ne sont pas capables de synthétiser la phéromone d'alarme. Les réponses contradictoires à cette phéromone (Minoli et al., 2013) ainsi que l'absence de sa synthèse chez les juvéniles soulèvent encore de nombreuses questions (Lazzari et al., 2013).

Comme chez tous les insectes hématophages, le dioxyde de carbone (CO_2), signalant la proximité des animaux à sang chaud, modifie le comportement de *Triatominae* (Barrozo and Lazzari, 2004a; Guerenstein and Hildebrand, 2008). Le CO_2 n'est pas le seul composant utilisé par les *Triatominae* pour repérer une ressource alimentaire. Le nonanal (Guerenstein and Guerin, 2001), l'acide isobutyrique et le 1-octène-3-ol retrouvés dans la sueur humaine (Barrozo and Lazzari, 2004b; Guerenstein and Guerin, 2001) et l'ammoniac retrouvé dans l'urine et la sueur des vertébrés (Taneja and Guerin, 1997) servent également de signaux olfactifs pour indiquer une source d'alimentation aux punaises hématophages. La communication chimique a donc un rôle très important dans la détection des sources alimentaires potentielles. Ainsi, la détection des odeurs émises par les hôtes stimule les *Triatominae* en recherche alimentaire et les incite à quitter leurs refuges. Cependant, leur réponse au CO_2 est modulée par un rythme circadien endogène (Barrozo and Lazzari, 2004a; Bodin et al., 2008), les conduisant à répondre uniquement pendant les premières heures de la nuit, où leur motivation à l'alimentation est également plus élevée (Lorenzo and Lazzari, 1998). La réponse des *Triatominae* à la présence d'hôtes dépend également de leur état physiologique, par exemple lorsqu'ils sont en période de

mue et de reproduction (Bodin et al., 2009a, 2009b). Par ailleurs, en fonction de son statut nutritionnel, une punaise répondra différemment à l'émission de CO₂. Si elle est à jeûn, ce signal sera considéré comme une source de nourriture potentielle. En revanche, si elle vient de s'alimenter, elle sera repoussée par l'émission de CO₂ considérant le signal comme un prédateur potentiel (Bodin et al., 2009a).

B. Identification des gènes chimiosensoriel chez les Triatominae

A ce jour, le système chimiosensoriel des Triatominae a été très peu étudié. Les études du sialotranscriptome de différents Triatominae ont cependant permis l'identification de quelques transcrits chimiosensoriels. Ainsi, Assumpção et al., (2008) ont annoté dans le sialotranscriptome de *T. infestans* deux OBPs partageant des similitudes avec celles des moustiques *An. gambiae* et *A. Aegypti*. Cependant, seules quatre cystéines sur les six conservées et caractéristiques des OBPs ont été retrouvées suggérant qu'il pourrait peut-être s'agir soit de CSPs soit d'OBPs Minus-C. Ces protéines de liaison retrouvées dans les glandes salivaires de *T. infestans* pourraient être impliquées dans la perception gustative (Assumpção et al., 2008). Dans le transcriptome du système digestif de *R. prolixus*, certains contigs ont été annotés comme OBPs (11 contigs) et CSPs (11 contigs) (Ribeiro et al., 2014). Ces OBPs et CSPs exprimées dans l'intestin de *Rhodnius*, ont probablement d'autres fonctions que l'olfaction. Elles pourraient éventuellement assurer le transport de nutriments ou de molécules impliquées dans la coordination des fonctions physiologiques de l'intestin (Ribeiro et al., 2014).

Le récent séquençage du génome de *Rhodnius prolixus* a permis l'annotation des différentes familles de gènes. Ainsi, 27 OBPs et 19 CSPs ont été annotées (Mesquita et al., 2015). Plusieurs OBPs et CSPs sont organisées en groupes localisés sur un même supercontig. On retrouve ainsi 3 groupes comprenant 4 à 9 OBPs réparties sur 3 supercontigs et un groupe de 6 CSPs sur un 4^{ème} supercontig. Les récepteurs ont également été annotés sur le génome de *R. prolixus* (Mesquita et al., 2015 ; Latorre-Estivalis et al., 2015a). Ainsi, 116 gènes ont été annotés comme ORs potentiels. Parmi ces 116 ORs annotés, 5 sont des pseudogènes et 16 sont incomplets (dont 7 du fait de scaffolds fragmentés dans l'assemblage du génome). Parmi ces ORs, un seul codant un corécepteur a pu être identifié. Un jeu de 28 gènes codant des GRs a été détecté chez *R. prolixus*. Seule une des grandes sous-familles de GRs connue chez les insectes est présente chez *R. prolixus*. Cette dernière serait impliquée dans la perception du fructose. En revanche, la sous-famille conservée des récepteurs des autres sucres ainsi que celle des récepteurs de CO₂ n'ont pas été retrouvées dans le génome de cette punaise hématophage. Par ailleurs, *R. prolixus* possède une sous-famille de GRs qui lui est propre et qui serait potentiellement impliquée dans la perception du goût amer. Trente trois IRs ont été identifiés dans le génome de *R. prolixus*, répartis en sept sous-familles ce qui diverge de la situation de la drosophile qui comporte seulement deux sous-familles (Mesquita et al., 2015 ; Latorre-Estivalis et al., 2015a) Trois corécepteurs ionotropiques ont peu être identifiés (Latorre-Estivalis et al.,

2015a). Cependant, en l'absence d'étude fonctionnelle, le rôle de ces différents gènes chimiosensoriels n'est, à l'heure actuelle, pas validé.

Les profils d'expression des ORs et des IRs ont été évalués sur les différents stades de *R. prolixus* (10ORs et 20IRs testés) et sur différents organes de mâles et de femelles susceptibles d'intervenir dans la chimioréception (7ORs et 14IRs testés) (Latorre-Estivalis et al., 2015a). Certains ORs et IRs sont exprimés dans tous les tissus analysés, mâles et femelles, tandis que d'autres sont spécifiques à un seul organe et/ou à un seul sexe. Le corécepteur olfactif et les trois corécepteurs ionotropiques ont été détectés dans tous les organes (hormis le rostre de la femelle, le tarse et les organes génitaux mâles pour l'IR8a) et dans tous les stades de développement (Franco et al., 2015; Latorre-Estivalis et al., 2015b). D'autres récepteurs ne sont jamais exprimés dans les antennes quel que soit le stade. D'autres sont exprimés dans tous les stades hormis l'embryon pour certains. Cependant, la plupart de ces récepteurs montrent une tendance à l'augmentation de leur expression au cours du développement de la punaise. Certains d'entre eux sont exprimés spécifiquement chez les adultes, suggérant leur implication dans la détection de phéromones sexuelles (Latorre Estivalis et al., 2015a; Latorre-Estivalis et al., 2015b).

La présence du corécepteur olfactif dans tous les stades de *Rhodnius* pourrait être expliquée par le fait que chaque stade nécessite la prise de repas sanguin, et donc la recherche d'hôtes. Or, il a été montré chez *Anopheles gambiae* que la prise de repas sanguin avait un effet inhibiteur sur l'expression du corécepteur olfactif et que la sous-expression de ce gène était corrélée à une réduction de la réponse aux odeurs humaines (Fox et al., 2001). Des résultats similaires ont été trouvés chez *Rhodnius* au stade adulte (Latorre Estivalis et al., 2015b). Dans cette étude, la prise de repas sanguin induit la diminution de l'expression de RproOcr chez les deux sexes ainsi que celle des corécepteurs ionotropiques RproIr8a et 25a uniquement chez la femelle. Cependant, Franco et al., (2015) ont trouvé des résultats contraires, les RproOcr ne présentant pas de différence d'expression en fonction de l'état de nutrition des punaises adultes. Des études complémentaires sont donc nécessaires pour départager ces observations contradictoires. En revanche, ces auteurs ont montré que l'inhibition de ce co-récepteur par RNAi diminuerait considérablement la capacité des punaises à trouver leur hôte ainsi que l'ingestion de sang. Les protéines et les lipides fournis par les repas sanguins étant essentiels pour la production d'œufs, l'inhibition du corécepteur olfactif diminue de moitié la fertilité des *Rhodnius*. On pourrait donc envisager d'utiliser le gène RproOrco comme cible de lutte antivectorielle (Franco et al., 2015).

Matériels et Méthodes

I. Echantillonnage

1. *Triatoma brasiliensis*

Les échantillons de *T. brasiliensis* ont été récoltés par nos partenaires brésiliens (Carlos Eduardo Almeida et Jane Costa, Fiocruz, Brésil) dans la zone biogéographique de Caatinga de la municipalité de Caico (dans l'Etat de Rio Grande do Norte, au Nord-Est du Brésil, Figure 8). Les emplacements géographiques et les distances ont été obtenus en utilisant un GPS TrackMaker (Figure 8). L'écotopie domicilié correspond aux habitations où des personnes vivent ou travaillent. Les insectes se trouvaient généralement dans les crevasses de murs de boue ou sous les lits. Les captures des individus domiciliés ont été effectuées pendant la journée à l'aide d'un répulsif délogeant les punaises. L'habitat péri-domicilié correspond à un rayon de 50 mètres autour des habitations, où sont élevés les animaux domestiques. Dans cet habitat, la plupart des punaises ont été capturées la nuit dans les poulaillers et dans les enclos de chèvres et de porcs. L'habitat sylvatique est une zone éloignée d'au moins 200 mètres de toute habitation. Les punaises sylvatiques retrouvées sur un même complexe rocheux (Almeida et al., 2002; Sarquis et al., 2010) sont considérées comme appartenant à la même population. Les captures sylvatiques ont été faites de nuit. Les spécimens adultes ont été identifiés en fonction de leurs caractéristiques taxonomiques et leur statut a été confirmé par analyse moléculaire (cytochrome b).

Figure 8 : Cartographie des points de collecte des échantillons de *T. brasiliensis* dans les environs de Caico, Rio Grande do Norte (cf Almeida, C.E., Faucher, L., Lavina, M., Costa, J., and Harry, M. (2016). Molecular eco-epidemiology of *Triatoma brasiliensis* using an individually based approach: Inferences on Chagas disease transmission links. Mol. Ecol)

2. Les 10 espèces du genre *Rhodnius*

A. *Rhodnius prolixus*

Nous avons utilisé des échantillons provenant de deux souches différentes. La première souche est maintenue au laboratoire de parasitologie, à l'Institut de Biologie de Neuchâtel en Suisse (pour l'échantillon mâle PMES et l'échantillon femelle PFES). La seconde souche (souche 077) est maintenue au laboratoire Insetário de Triatominae da Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP) - Araraquara and Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos (pour l'échantillon mâle PM'EF et l'échantillon femelle PFEF) (Tableau 2).

B. *Rhodnius robustus*

Des échantillons de *R. robustus* sauvages ont été récoltés au nord de la Guyane Française dans la région de Cayenne (RMG1, RFG1 et RFG2). Ils ont été échantillonnés avec des pièges lumineux à Macouria ou retrouvés dans les maisons à Montjoly lors de visites

sporadiques (Figure 9). Dans ce dernier cas, les punaises entrent dans les maisons pour se nourrir mais ne s’y installent pas pour former des colonies. De ce fait, ces punaises ne sont pas considérées comme domiciliées et gardent leur statut de population sylvatique. Afin de se rapprocher de la condition domiciliée, nous avons utilisé dans notre étude des échantillons (RMEF1 et RFEF1 ; Tableau 2) provenant d’un élevage d’un laboratoire brésilien (Insetário de Triatominae da Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP) – Araraquara).

Figure 9 : Cartographie des points de collecte des échantillons de *R.robustus* dans les environs de Cayenne en Guyane Française (cf Macouria et Montjoly, Guyane Française, Google earth 2015)

C. Les 8 autres espèces *Rhodnius*

Huit autres espèces de *Rhodnius* ont été utilisées dans cette étude provenant du même laboratoire brésilien (Tableau 2). Cet échantillonnage regroupe des espèces domiciliées (*R. pallescens*), en voie de domiciliation (*R. neglectus*, *R. nasutus*) et sylvatiques (*R. sp.*, *R. pictipes*, *R. montenegrensis*, *R. brethesi*, *R. colombiensis*). Notons que l’espèce *R. sp* non encore décrite est affine de l’espèce *R. neglectus*.

II. Analyse de données RNAseq

1. Préparation des données

Plusieurs individus ont été regroupés par stade, par sexe et par population, à l’exception de l’échantillon SFF qui correspond à une seule femelle sylvatique (Tableau 1). Pour cibler les gènes impliqués dans le système chimiosensoriel, l’ARN a été extrait à partir des antennes et des rostrs des punaises. Pour certains échantillons, la tête a été ajoutée au protocole d’extraction afin de recueillir suffisamment d’ARN pour le séquençage et augmenter la représentativité du transcriptome.

Deux types de séquençage ont été réalisés : séquençage 454 et Illumina par les plateformes LGC et Imagif. Le séquençage en technologie 454 nécessite une quantité très importante d’ARN. Ainsi, les aliquotes d’ARN de tous les échantillons de *T. brasiliensis* adultes extraits (excepté l’échantillon SFF) ont été rassemblées pour la construction d’une librairie normalisée d’ADNc. Celle-ci a été séquencée en utilisant les technologies 454 Roche GS FLX. Le même procédé a été réalisé pour *R. prolixus*. En revanche, pour le séquençage Illumina, le séquençage a été réalisé échantillon par échantillon. Lorsque la quantité d’ARN était suffisante, nous avons réalisé des réplicats techniques. Ainsi quatorze librairies indépendantes ont été séquencées à partir de 8 échantillons de *T. brasiliensis* avec la méthode Illumina en single-read de 100 pb. Les librairies de l’échantillon adulte SFF de *T. brasiliensis* et

des différentes espèces de *Rhodnius* ont été séquencées en lectures paired-end (2x100pb) avec la technologie Illumina et une déplétion d'ARN ribosomiques.

Echantillon	Espèce	Sexe	Population	Environnement	Organes	Nbre indiv	Séquençage	Plateforme
DFB	<i>T. brasiliensis</i>	F	B	Domicilié	ant + ro	4	Illumina single	LGC
DMB1	<i>T. brasiliensis</i>	M	B	Domicilié	ant + ro	5	Illumina single	LGC
DMB2	<i>T. brasiliensis</i>	M	B	Domicilié	ant + ro	5	Illumina single	LGC
PFD1	<i>T. brasiliensis</i>	F	D	Péri-domicilié	ant + ro	6	Illumina single	LGC
PFD2	<i>T. brasiliensis</i>	F	D	Péri-domicilié	ant + ro	6	Illumina single	LGC
PMD	<i>T. brasiliensis</i>	M	D	Péri-domicilié	ant + ro	11	Illumina single	LGC
SFA1	<i>T. brasiliensis</i>	F	A	Sylvatique	ant + ro	8	Illumina single	LGC
SFA2	<i>T. brasiliensis</i>	F	A	Sylvatique	ant + ro	8	Illumina single	LGC
SMA1	<i>T. brasiliensis</i>	M	A	Sylvatique	ant + ro	10	Illumina single	LGC
SMA2	<i>T. brasiliensis</i>	M	A	Sylvatique	ant + ro	10	Illumina single	LGC
SFC1	<i>T. brasiliensis</i>	F	C	Sylvatique	ant + ro	4	Illumina single	LGC
SFC2	<i>T. brasiliensis</i>	F	C	Sylvatique	ant + ro	4	Illumina single	LGC
SMC1	<i>T. brasiliensis</i>	M	C	Sylvatique	ant + ro	15	Illumina single	LGC
SMC2	<i>T. brasiliensis</i>	M	C	Sylvatique	ant + ro	15	Illumina single	LGC
Total pop (Tbra454)	<i>T. brasiliensis</i>	F+M	B+D+A+C	Domicilié Péri-domicilié Sylvatique	ant + ro	63	454	LGC
SFF	<i>T. brasiliensis</i>	F	F	Sylvatique	ant + ro + tête	1	Illumina paired	ImaGIF

Tableau 1 : Récapitulatif des échantillons de *Triatoma brasiliensis* séquencés en haut débit.

Le sexe des échantillons est indiqué par F (Femelle) ou M (Mâle). Les échantillons proviennent de populations domiciliées (B), péri-domiciliées (D) ou sylvatiques (F, A ou C). L'ARN a été extrait à partir des organes chimiosensoriels antennes et rostre (ant+ro) ou à partir de la tête entière, organes chimiosensoriels compris (ant+ro+tête). Le nombre d'individus par échantillon est indiqué (Nbr indiv) ainsi que la technologie de séquençage utilisée et la plateforme de séquençage qui l'a réalisée.

Espèce	Echantillon ADN	Echantillon ARN (Nbr indiv / Sexe/ organes)	Distribution	Origine de l'échantillon	Habitat naturel
<i>Rhodnius pallescens</i>	I-R1J	1TF (3F / ant+ro)	Belize, Colombie, Costa Rica, Panama, Venezuela, Nicaragua	Laboratoire	Sylvatique et Domicilié ; principal vecteur au Panama
<i>Rhodnius sp</i>	Q-R3WT	3TF (3F / ant+ro)	Brésil	Laboratoire (Brésil, état de Sao Paulo 58)	Sylvatique
<i>Rhodnius neglectus</i>	R5WU	5TF (3F / ant+ro)	Brésil	Laboratoire (Brésil, état de Minas 61 Gerais)	Sylvatique, domiciliation en cours
<i>Rhodnius pictipes</i>	R-R6WV	6TF 3(F / ant+ro)	Belize, Bolivie, Brésil, Colombie, Guyane, Guyane Française, Pérou, Suriname, Venezuela, Equateur, Trinidad	Laboratoire (Brésil, état de Para)	Sylvatique, sporadiquement domicilié
<i>Rhodnius nasutus</i>	S-R7WU	7TF (3F / ant+ro)	Brésil	Laboratoire (Argentine)	Sylvatique, domiciliation en cours
<i>Rhodnius montenegrensis</i>	E-R8F	8TF (3F / ant+ro)	Brésil	Laboratoire (Monténégro)	Sylvatique
<i>Rhodnius brethesi</i>	T-R9WX	9TF (3F / ant+ro)	Brésil, Colombie, Venezuela	Laboratoire (Amazonie)	Sylvatique, sporadiquement domicilié
<i>Rhodnius colombiensis</i>	A-R10B	10TF (3F / ant+ro)	Colombie	Laboratoire (Colombie)	Sylvatique, sporadiquement domicilié
<i>Rhodnius prolixus</i>	Génome vectorbase	PFEF (18F / ant+ro) PMEF (18M / ant+ro) 454 (28F+28M / ant+ro+tête) PMES (5M / ant+ro+tête)	Brésil, Colombie, Costa Rica, Salvador, Equateur, Mexique, Guatemala, Guyane, Guyane Française, Honduras, Nicaragua, Panama, Suriname, Trinidad, Venezuela	Laboratoire Fiocruz	Domicilié, principal vecteur en Amérique centrale
		Laboratoire Suisse			
<i>Rhodnius robustus</i>	O-ROBR	RFEF (21F / ant+ro) RMEF (19M / ant+ro)	Brésil, Colombie, Guyane Française, Pérou, Venezuela, Bolivie, Equateur	Laboratoire (Pérou)	Sylvatique, sporadiquement domicilié
	P-ROB5S	RFG1 (1F / ant+ro + tête) RFG2 (2F / ant+ro) RMG (4M / ant+ro)		Guyane Fr	

Tableau 2 : Récapitulatif des échantillons de *Rhodnius* (ADN et ARN) séquencés en haut débit. L'espèce, sa distribution et la provenance de l'échantillon sont indiquées. Les échantillons d'ARN ont été extraits à partir des organes chimiosensoriels antennes et rostre (ant+ro) ou à partir de la tête entière, organes chimiosensoriels compris (ant+ro+tête). Les échantillons d'ARN sont issus du regroupement de plusieurs individus femelles (F) et/ou mâles (M) (le nombre d'individus par échantillon est indiqué devant le sexe).

Lors du séquençage, les premières bases d'une lecture sont très souvent séquencées avec beaucoup de fiabilité, mais plus on avance dans la séquence, et moins la qualité est bonne. Or, des erreurs de séquençage sur des lectures de mauvaise qualité vont complexifier considérablement le travail des logiciels d'assemblage. Les lectures ont donc été préalablement nettoyées selon leur qualité grâce au logiciel Prinseq (Schmieder and Edwards, 2011) après élimination des adaptateurs avec Cutadapt (Martin, 2011). La qualité des séquences ainsi obtenue a été validée en utilisant le logiciel FastQC (Marchant et al., 2015a, 2015b).

Le génome de *T. brasiliensis* n'ayant pas encore été séquencé, la construction d'une référence était indispensable avant toute analyse de comparaison. L'assemblage *de novo* consiste à reconstituer un transcriptome sans référence à partir d'un ensemble de lectures. Les transcrits sont reconstitués par chevauchement de séquences des lectures. Dans le but de réaliser un transcriptome le mieux assemblé et le plus complet possible, plusieurs programmes d'assemblage et plusieurs workflows ont été utilisés et comparés à partir des lectures obtenues avec les échantillons de *T. brasiliensis*.

Dans une première étape, les données 454 ont été assemblées séparément des données Illumina avec Newbler, MIRA (Chevreux et al., 2004) et CAP3 (Huang and Madan, 1999). Par ailleurs, les lectures d'Illumina ont été assemblées à l'aide de Trinity (Grabherr et al., 2011). Dans une deuxième étape, plusieurs essais d'assemblage combinés des données Illumina et des données 454 ont été réalisés dans le but d'augmenter la longueur des contigs. Ces différents tests d'assemblage *de novo* ont également été réalisés pour les échantillons de *R. prolixus* et de *R. robustus*. En absence de données 454 pour les autres espèces de *Rhodnius*, seul l'assemblage Trinity a été réalisé.

Le génome de *R. prolixus* ayant été séquencé, différents assemblages basés sur cette référence ont également été réalisés pour le transcriptome de cette espèce et pour celui de son espèce sœur *R. robustus*. Pour cela, les lectures nettoyées des différents échantillons ont été alignées sur le génome de *R. prolixus* grâce au logiciel Tophat (Trapnell et al., 2009). Ce logiciel permet de détecter les jonctions d'épissage et ainsi d'aligner les lectures à cheval sur deux exons réunis après épissage de l'ARN. Les lectures ont ensuite été assemblées avec le logiciel Cufflinks (Trapnell et al., 2010). Différents tests ont été réalisés avec ou sans l'apport des informations d'annotation du génome pour l'alignement et/ou l'assemblage.

Plusieurs critères d'évaluation des transcriptomes ont été utilisés pour choisir le meilleur assemblage : le nombre de contigs, le N50 (longueur du contig pour laquelle les séquences les plus grandes couvrent 50 % de la longueur totale de l'assemblage) ainsi que la complétude des transcriptomes.

A partir des transcriptomes de *T. brasiliensis* et *R. prolixus* considérés comme les mieux assemblés et à partir de l'ensemble des assemblages Trinity des différentes espèces de *Rhodnius*, les contigs codant les OPBs et CSPs ont été annotés. Pour cela, une base de données protéique d'OBPs et de CSPs de différents insectes a été alignée sur les transcriptomes à l'aide du programme Blastx. Les contigs annotés dans cette première phase ont été validés par une seconde phase de Blast sur la base de données non-redondantes (NR) des protéines. Puis après traduction de ces contigs en protéines, la présence d'un signal peptidique et d'hélices alpha ainsi que la position caractéristique de 6 et 4 cystéines ont été vérifiées pour les OBPs et les CSPs respectivement.

2. Alignement, comptage et analyse statistique

Afin d'évaluer l'expression de chaque contig pour les différentes populations de *T. brasiliensis*, chaque lecture de chaque échantillon a été attribuée à un contig du transcriptome de référence. Pour cela, les lectures ont été alignées sur le transcriptome avec le logiciel BWA (Li and Dublin, 2009). La présence d'isoformes dans nos jeux de données complexifie cette étape. Pour limiter cette problématique, seuls les plus grands contigs par groupe d'isoformes définis par Trinity ont été sélectionnés pour réaliser l'étape d'alignement. Par ailleurs, les lectures alignées à plusieurs endroits ont été éliminées du comptage final pour ne pas biaiser les analyses d'expression. Le nombre de reads alignés sur chaque contig a été évalué pour chaque échantillon grâce à l'outil Samtools. Ces données de comptage ont été utilisées pour les tests statistiques d'expression différentielle. Les différentes conditions comparées étaient celles de l'environnement et du sexe des individus échantillonnés. Différents packages R ont été utilisés pour réaliser ces analyses: DESeq2 (Love et al., 2014) et EdgeR (Robinson et al., 2010). Les résultats obtenus ont été comparés et seuls les contigs retrouvés significativement différentiellement exprimés avec les deux packages ont été pris en compte. Une analyse de

regroupement des contigs co-exprimés a été réalisée à l'aide du logiciel HTSCluster (Rau et al., 2015).

3. Corrélation entre le nombre de CSPs et OBPs définies chez les *Rhodnius* et leur capacité de domiciliation

Afin de tester la corrélation entre le type d'habitat des espèces de *Rhodnius* et le nombre de CSPs/OBPs, il est nécessaire de prendre en considération les liens phylogénétiques qui relient les différentes espèces entre elles. Nous avons d'abord établi une phylogénie robuste par vraisemblance à partir de 43 gènes nucléaires homologues issus des échantillons de *Rhodnius*. Ensuite, les OBPs et les CSPs ont été annotées dans les génomes des 10 espèces de *Rhodnius* à l'aide des OBPs et des CSPs connues chez *R. prolixus* et de l'annotation des transcriptomes. Puis, un test de contraste indépendant de la phylogénie (« Phylogenetic Independent Contrast » ou « PIC ») a été réalisé (Felsenstein, 1985). Cette méthode consiste à déterminer la variance de toutes les branches de l'arbre, pour ensuite étudier la corrélation entre le nombre de gènes chimiosensoriels et l'habitat pondérée par les liens phylogéniques.

*Partie I : Analyse du transcriptome
chimiosensoriel de Triatoma brasiliensis*

Chapitre 1 : Elaboration et annotation d'un transcriptome de référence

Résumé en français

Introduction : Le séquençage à haut débit offre de nouvelles perspectives d'étude en écologie, notamment sur les espèces non modèles. Le RNAseq en particulier, offre des informations cruciales sur la régulation de l'expression de l'ensemble des gènes, ce qui nous aide à mieux comprendre les liens reliant phénotype et génotype. Ainsi, la comparaison de transcriptomes d'individus provenant d'environnements différents peut nous renseigner sur les mécanismes d'adaptation des organismes à un milieu *via* la régulation de leurs gènes. Cependant, la comparaison de données transcriptomiques nécessite une référence sur laquelle seront cartographiées les lectures résultant du séquençage à haut débit des transcriptomes étudiés. En l'absence de génome de référence, il faudra construire un transcriptome de référence *de novo*. La qualité de cette référence va dépendre du choix de la procédure d'échantillonnage, de la technologie de séquençage et des stratégies d'assemblage. Afin d'élaborer un transcriptome de référence fiable pour notre espèce modèle *Triatoma brasiliensis*, le principal vecteur de la maladie de Chagas au nord du Brésil, différents protocoles d'assemblage *de novo* ont été testés en utilisant divers ensembles de données et de logiciels.

Matériels et Méthodes : L'ARN de *T. brasiliensis* a été extrait à partir des organes chimiosensoriels (rostre et antennes) d'un groupe de plusieurs insectes ou d'une seule punaise (dans ce cas, la tête a été ajoutée au protocole d'extraction). Trois technologies de séquençage ont été utilisées : le 454, l'Illumina single-end, et l'Illumina paired-end. Après nettoyage des lectures, plusieurs assembleurs ont été testés : Newbler, CAP3 et/ou MIRA ont été utilisés pour les lectures 454 alors que Trinity a été utilisé pour les lectures Illumina. Dix protocoles d'assemblage ont été testés en utilisant ces programmes séparément ou en les combinant. Pour comparer les résultats obtenus, plusieurs critères ont été utilisés : la taille du N50, le nombre de contigs, le pourcentage estimé de contigs chimériques et l'estimation par CEGMA de la complétude du transcriptome.

Résultats : La librairie 454 a fourni 555 854 lectures, les quatorze librairies Illumina single-end construites ont donné près de 360 millions de lectures et la librairie paired-end a généré 46 millions de lectures. Les différents assemblages ont généré de 15 825 à 365 717 contigs dont 0,04 à 1,05% sont potentiellement chimériques. Les N50 de ces différents transcriptomes s'étendent de 313 à 1 393bp. La complétude des assemblages est comprise entre 25 et 79,44%. Parmi ces assemblages, celui qui présente le meilleur compromis entre les différents critères d'estimation de la qualité d'assemblage a été réalisé à partir de plusieurs assembleurs combinés et basé sur l'utilisation à la fois de lectures Illumina paired-end séquencées à partir d'un seul individu et de données 454 dont les erreurs de séquençage d'homopolymères ont été corrigées. Il comporte 57 657 contigs dont moins de 1% sont potentiellement chimériques. La complétude de ce transcriptome est estimée à 79.44%.

Conclusion : L'extraction d'ARN provenant d'un seul individu facilite le travail des assembleurs en réduisant le taux de polymorphisme de la librairie. L'assemblage hybride qui fusionne à la fois des lectures longues (454) et des lectures courtes mais avec une couverture plus importante (Illumina paired-end) semble être une bonne stratégie. Enfin, l'utilisation de plusieurs assembleurs permet de combiner les avantages spécifiques de chacun et d'améliorer le transcriptome final.

Article 1: De novo transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease (*Genetica*, April 2015, Volume 143, Issue 2, pp 225-239).

A. Marchant^{*1,2}, F. Mougél^{1,2}, C. Almeida³, E. Jacquin-Joly⁴, J. Costa⁵, M. Harry ^{*1,2}

(1) Laboratoire Evolution, Génomes et Spéciation LEGS UPR 9034 - CNRS - avenue de la Terrasse, Bâtiment 13, BP1 - 91198 Gif-sur-Yvette - France

(2) Université Paris Sud - Orsay - France

(3) Departamento de Ciências Biológicas, Faculdade de Ciências Farmacêuticas (UNESP), Araraquara, SP; UNESP – Brazil

(4) INRA, UMR 1392, Institut d'Ecologie et des Sciences de l'Environnement de Paris, - Route de Saint Cyr, 78026 Versailles - France

(5) Laboratório de Biodiversidade Entomológica; Instituto Oswaldo Cruz, Fiocruz - Rio de Janeiro - Brazil

Abstract

High throughput sequencing (HTS) provides new research opportunities for work on non-model organisms, such as differential expression studies between populations exposed to different environmental conditions. However, such transcriptomic studies first require the production of a reference assembly. The choice of sampling procedure, sequencing strategy and assembly workflow is crucial. To develop a reliable reference transcriptome for *Triatoma brasiliensis*, the major Chagas disease vector in Northeastern Brazil, different *de novo* assembly protocols were generated using various datasets and software. Both 454 and Illumina sequencing technologies were applied on RNA extracted from antennae and mouthparts from single or pooled individuals. The 454 library yielded 278 Mb. Fifteen Illumina libraries were constructed and yielded nearly 360 million RNA-seq single reads and 46 million RNA-seq paired-end reads for nearly 45 Gb. For the 454 reads, we used three assemblers, Newbler, CAP3 and/or MIRA and for the Illumina reads, the Trinity assembler. Ten assembly workflows were compared using these programs separately or in combination. To compare the assemblies obtained, quantitative and qualitative criteria were used, including contig length, N50, contig number and the percentage of chimeric contigs. Completeness of the assemblies was estimated using the CEGMA pipeline. The best assembly (57,657 contigs, completeness of 80%, less than 1% chimeric contigs) was a hybrid assembly leading to recommend the use of i) a single individual with large representation of biological tissues, ii) merging both long reads and short paired-end Illumina reads, ii) several assemblers in order to combine the specific advantages of each.

Keywords

HTS, *de novo* assembly, non-model organisms, chimeric contigs, Chagas disease vector, *Triatoma*

Introduction

HTS has opened considerable opportunities in the field of genomics for the study of a number of organisms (Martin and Wang 2011; Stapley et al. 2010; Werner 2010). Through massive sequencing approaches, differential expression studies, candidate gene searches, gene ontology annotation, molecular marker development, alternative splicing, genetic variation characterization and SNP development can now be done in non-model species (Ekblom and Galindo 2011). Such information could be of interest for insect vectors in order to improve control strategies. However, comparative studies first require a reference assembly that can either be a genome or a transcriptome.

Concerning the bloodsucking bugs transmitting to human the parasite *Trypanosoma cruzi*, responsible of the Chagas disease, endemic to Latin America, only one genome that of *Rhodnius prolixus*, was recently sequenced and assembled (www.vectorbase.org) was the main vector of Chagas disease in Northern Latin America and Central America before its recent eradication (Hashimoto and Schofield 2012). But, in Brazil, since the recently eradication of the main vector, *Triatoma infestans*, by massive campaigns of chemical control (Dias et al., 2002), health authorities are now concerned about *Triatoma brasiliensis*, another vector increasingly reported in houses and peridomestic environment (Almeida et al., 2008; Borges et al., 2005; Costa, 1999; Costa et al., 2003, 1998). Comparative transcriptomic study based on the use of the newly available high throughput sequencing (HTS) technologies would allow deciphering the molecular bases of the domiciliation process.

Generating such a reference from a non-model species can be complex since assembling *de novo* the considerable amount of reads generated from HTS is a difficult puzzle. Assembling transcriptomic rather than genomic data may appear easier, facilitating further differential expression studies, as non-coding sequences are absent and the amount of information reduced (Paszkiwicz and Studholme 2010). Nevertheless, alternative splicing, gene duplication or intrinsic polymorphism of biological sample may complicate transcriptome assembly (Sammeth 2009). HTS has different characteristics depending on the technology, which may affect the assembly process. Among the conventional techniques, the substantial length (approximately 342 bp) of 454 reads facilitates contig assembly, but this technique also generates a high frequency of homopolymer repetition errors. Furthermore, read coverage of transcripts remains low with this technique because of the small amount of reads produced: typically around hundreds of thousands of reads (Ekblom and Galindo, 2011). As an alternative, Illumina technology sequencing, which may be done on single or paired-end reads, produces a very large amount of reads (typically a billion), generating a high coverage depth. However, the small size of the sequences (about 100 bp) complicates the assembly. Both technologies can be exploited for their respective advantages by combining long reads from 454 with high coverage depth from Illumina data (Martin and Wang 2011).

The choice of assembly software can be complex. Indeed, with the development of HTS, the publication of various *de novo* transcriptome assemblers, based on more or less similar

approaches, has boomed. For example, many software programs are available for long reads typically obtained with 454 technology, including MIRA (Chevreux et al. 2004), CAP3 (Huang and Madan 1999), Newbler (Knudsen et al. 2007), NGen (www.vectorbase.org), or CLCbio (<http://www.clcbio.com>). Among these, three programs are highly recommended: MIRA, which gives very reliable output according to (Mundry et al., 2012; Ren et al., 2012); CAP3 that was suggested as a robust assembling software by Cahais et al. (2012); and the commercial program Newbler, often furnished by the sequencing platform used and recommended by Ren et al. (2012). MIRA is the only program that combines the iterative approach and the all-in-one-step approach. The first step of iterative approach produces an assembly with very stringent parameters (particularly with regard to the quality criteria). Following steps continue with less stringent criteria until the assembly is finished. The all-in-one-step approach starts with all data whatever its quality and generates a consensus using best quality bases as reference. This way, MIRA uses iterative multipass strategies for reads targeted as high or acceptable quality for the assembly. At this step, it detects SNP during the assembly and sorts them by type to reduce assembly mistakes. In a second stage, a fallback strategy uses reads with lower quality if necessary (Chevreux et al. 2004). CAP3 is based on a multiple alignment strategy, taking into account the base quality (Huang and Madan 1999). Newbler starts by identifying overlaps between reads and next builds multiple alignments from overlapping reads to generate preliminary contigs. The final step consists in creating a contig graph to connect contigs together from the edges (Knudsen et al., 2007). To assemble short reads typically produced by Illumina technology, various free software are available, among which some are specifically devoted to transcriptomic data: Trans-Abyss (Robertson et al. 2010), Trinity (Grabherr et al., 2011), Oases (Schulz et al. 2012), Rnnotator (Martin et al. 2010), Multiple-K (Surget-Groba and Montoya-Burgos 2010) or SOAPdenovo-Trans (Xie et al. 2013). Trinity, which was especially designed for transcriptome assembly and offers several advantages: it can solve 99% of the initial sequencing errors, detects isoforms from the alternative splicing or transcripts recently duplicated and identifies allelic variants while building complete transcripts (Grabherr et al. 2011).

To determine the most appropriate assemblers, various comparative studies have been conducted on biological datasets or on simulated coding DNA sequences (Cahais et al. 2012; Mundry et al. 2012; Ren et al. 2012; Vijay et al. 2013; Zhao et al. 2011). Comparisons were performed based on three groups of criteria. The "computer" criteria include the amount of RAM required and the run time. The quantitative criteria comprise the number of contigs, the total number of bases assembled, the average length of contigs, the N50 (length for which the collection of all contigs of that length or longer contains at least half of the total of the lengths of the contigs), the number of contigs higher than 1 kb and the number of mapped reads on the final assembly. The qualitative criteria include the completeness of the transcriptome and the presence of chimeric contigs, both parameters being obtained by comparison with a reference genome. The programs considered as the most effective differ between comparative

studies depending on datasets and on the weight attributed to different evaluation criteria. Based on a simulated dataset, Vijay et al. (2013) recommended SOAPdenovo, which provides non chimeric, but fragmented, transcriptome, especially when there are a large number of reads and high polymorphism in the dataset (Vijay et al. 2013; Zhao et al. 2011). Among assemblers based on the single-k-mer strategy, Zhao et al. (2011) identified Trinity as the best assembler for Illumina datasets because of the greater number of contigs in the assemblies that can be aligned to the reference genome. For 454 reads from transcriptomes, Ren et al. (2012) recommended MIRA, while Mundry et al. (2012) recommended Newbler and MIRA both showing contrasting specificities. MIRA generated a fragmented and redundant assembly while Newbler built a more complete restoration of full-length transcripts, although it contained large chimeric contigs. Cahais et al. (2012) developed a more complex workflow including data from 454 and Illumina using Abyss and CAP3 assemblers.

All these comparison studies are based on data generated from well-known species. By contrast, without a reference genome, the number of genes and alternative splicing are unknown. Therefore, it is difficult to estimate the expected number of transcripts, the completeness, quality and validity of the assembly or to detect chimeric contigs (Cahais et al. 2012). Definition of assessment methods should help in the choice of assembling strategy. Furthermore, all studies, except Cahais et al. (2012) study, compared assemblies generated from a single sequencing source (Illumina or 454) while both sources have complementary properties that could be combined.

To date, no massive genomic and transcriptomic data are available for *T. brasiliensis* except the production of a salivary gland cDNA library (Santos et al., 2007). The first step of such an approach is then to produce a reliable reference for *T. brasiliensis*. In this study, we applied HTS to RNA extracted from antennae, rostrum and head of *T. brasiliensis* males and females collected in various environments: sylvatic, domiciliary or peridomiciliary, to generate an optimized chemosensory transcriptome. The generated assembly is to be used as a reference for further comparative transcriptomics. It is known that chemosensation allows bugs to interact with their environment and influences their behavior, thus participating in insect adaptation to new environments and new host choice, playing an important role in anthropogenic adaptation. Studying the expression of the chemosensory transcriptomes from various populations of Triatominae would be an innovative way to address the issue of domiciliation.

For the purpose of getting the best assembly, several types of data and assemblers were tested and compared. Two sequencing technologies were used (454 and Illumina in single or paired-end) with samples from a single individual or pooled individuals. Ten assembly workflows were compared and characterized according to a set of qualitative and quantitative criteria. The hybrid assembly using Illumina paired-end and 454 corrected data in hierarchical workflow using several assemblers was found to optimize both the number of contigs, contig size and the completeness of transcriptome.

Materials and methods

Sampling

Samples of *T. brasiliensis* were collected in March 2011, in the Caatinga biogeographic zone of the municipality of Caico in the north of Brazil. Five populations were collected in sylvatic, domestic and peridomestic habitats (Table 1).

RNA extraction and sequencing

Several individuals were pooled by sex and population, except for the SFF sample which consisted of one sylvatic female (Table 1). To target genes involved in the chemosensory system, RNA was extracted from the antennae and rostrum of blood-sucking bugs using the TRIzol® Reagent kit (Invitrogen, Carlsbad, CA, USA). Aliquots of all extraction samples were pooled and the mix was used as template for the construction of a concatamerized, normalized cDNA library prepared and sequenced using 454 Roche GS FLX Titanium (½ Pico Titer Plate, , Tbra454, Table 1) by LGC Genomics GmbH (Berlin, Germany). In addition, fourteen libraries were constructed using TruSeq RNA Kit 2010 from Illumina for each RNA sample independently and sequenced with Illumina HiSeq2000 method (<http://support.illumina.com/>) in single-end reads of 100 bp by LGC Genomics platform. One library (SFF sample) was prepared using ScriptSeq™ Complete Kit (Human/Mouse/Rat)—Low Input of Epicentre (<http://www.epibio.com/>) and sequenced by Imagif (CNRS, Gif-sur-Yvette, France) in paired-end reads (2x100pb) with Illumina HiSeq1000, using TruSeq PE Cluster Kit v3-cBot-HS and TruSeq SBS Kit v3-HS kits.

Sample (assembly)	Number of individuals (RNA quantity/μl)	Organs	Sequencing	Number of initial reads	Number of post-cleaned process reads	Deconseq (%)	Length of reads	Contig number	Total number of bases	Length distribution (bp)				CEGMA complete (%)	CEGMA partial (%)	Potential chimeric contig nb (%)
										max	median	average	N50			
DFB (6-a1)	4 (210ng)	ant + ro	Illumina single	16,589,869	6,364,934	1.46	100 bp	8,912	2,454,964	3136	242	275	260	0	1.21	0.01
DMB1 (6-a2)	5 (810ng)	ant + ro	Illumina single	34,514,587	4,197,000	6.85	100 bp	14,852	4,581,021	3,884	263	308	298	3.23	16.94	0.03
DMB2 (6-a3)	5 (810ng)	ant + ro	Illumina single	36,948,992	5,403,327	3.78	100 bp	11,426	3,452,353	2,735	260	302	291	1	4.84	0.03
PFD1 (6-a4)	6 (830ng)	ant + ro	Illumina single	19,080,671	5,043,986	0.41	100 bp	24,142	7,901,016	2,744	275	327	322	4	18.95	0.02
PFD2 (6-a5)	6 (830ng)	ant + ro	Illumina single	18,163,626	3,639,582	1.35	100 bp	18,088	5,775,897	2,736	269	319	313	2	14.52	0.01
PMD (6-a6)	11 (28ng)	ant + ro	Illumina single	5,428,910	1,353,558	0.62	100 bp	3,544	1,049,958	1,757	257	296	284	0	1.21	0
SFA1 (6-a7)	8 (830ng)	ant + ro	Illumina single	24,892,774	6,274,583	1.45	100 bp	24,037	7,714,974	5,413	269	320	314	6	23.39	0.02
SFA2 (6-a8)	8 (830ng)	ant + ro	Illumina single	16,857,756	4,864,800	2.55	100 bp	16,478	5,017,124	3,488	260	304	293	2	14.92	0.02
SMA1 (6-a9)	10 (570ng)	ant + ro	Illumina single	53,062,536	5,829,446	0.27	100 bp	20,569	6,715,565	2,856	272	326	320	6	24.19	0.02
SMA2 (6-a10)	10 (570ng)	ant + ro	Illumina single	28,396,196	5,976,442	1.25	100 bp	22,636	7,314,022	2,578	272	323	316	6	22.18	0.02
SFC1 (6-a11)	4 (540ng)	ant + ro	Illumina single	38,431,926	9,497,163	2.65	100 bp	66,830	22,901,564	17,225	259	342	319	37	56.05	0.07
SFC2 (6-a12)	4 (540ng)	ant + ro	Illumina single	36,838,494	7,382,370	1.52	100 bp	10,682	3,041,341	4,768	246	284	270	1	7.66	0.01
SMC1 (6-a13)	15 (890ng)	ant + ro	Illumina single	22,712,152	5,103,875	5.32	100 bp	14,698	4,638,439	2,471	267	315	306	1	13.31	0.01
SMC2 (6-a14)	15 (890ng)	ant + ro	Illumina single	7,750,706	1,492,438	2.01	100 bp	8,157	2,515,977	1,830	263	308	298	2	4.84	0.01
Total pop (6-a)	63	ant + ro	Illumina single	359,669,195	74,386,931	NA	100 bp	256,799	85,972,145	26,078	261	334	313	60	83.47	0.05
Total pop (6-b)	63	ant + ro	Illumina single	359,669,195	59,659,197	NA	100 bp	136,755	45,898,586	26,182	258	335	313	43.15	71.37	0.07
SFF paired (6-c)	1 (128ng)	ant + ro + head	Illumina paired	46,185,332 R1 46,185,332 R2	15,762,629 R1 15,762,629 R2	NA	100 bp	365,717	179,169,855	47,145	301	490	548	81.45	93.95	0.08
SFF single (6-c')	1 (128ng)	ant + ro + head	Illumina single simulated	46,185,332	46,051,230	NA	100 bp	298,132	153,921,537	26,026	305	516	628	79.44	89.52	0.1
Total pop (Tbra454)	63	ant + ro + head	454	555,854	498,567	NA	average: 343 bp									

Table 1: *T. brasiliensis* samples from which transcriptomes were sequenced and characteristics of assemblies generated with Trinity

The first letter of each sample name indicates its environment: (D for domestic, P for peridomestic and S for sylvatic); the second letter indicates the gender (M for male and F for female); and the third letter is a population index. The number when present (1 and 2) corresponds to technical replicates (no replicates were performed for DFB, PMD and SFF). “6” corresponds to Trinity assembly of reads preprocessed with cleaned-1 (-a); with more stringent preprocessing cleaned-2 (-b) and with the single individual sample (-c). RNAs were extracted from antennae (ant) and rostrum (ro) for Illumina single-reads. RNA from head was added for Illumina paired-end.

Preprocessing

Illumina single-end reads were cleaned with Cutadapt (Martin 2011) and Prinseq (Schmieder and Edwards 2011). Two sets of thresholds were applied (the gentle "cleaned-1" and very stringent "cleaned-2") resulting in the deletion of: i) indefinite bases, ii) the poly A tails (using parameters trim-tail-left 5 for both cleanings; trim-tail-right 5 for cleaned-1 and trim-tail-right 8 for cleaned-2), iii) the small sequences (minimum size of 20 for cleaned-1, 40 for cleaned-2), iv) exact duplicated sequences resulting from PCR (for both cleanings), v), the low quality ends of the sequences (minimum grade of 20 for cleaned-1, and 28 for cleaned-2), and vi) the low complexity sequences (both cleanings). The cleaned-1 protocol was also applied for 454 reads and Illumina paired-end reads. For samples sequenced in paired-end, reads that became singlets after cleaning were set aside before assembly.

The quality of the sequences was analyzed using FastQC (Supplementary data S1; <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Part of possible contamination of cleaned-1 reads was evaluated for each single-read sample using Deconseq version 0.4.1 (Schmieder and Edwards 2011b). We used several databases: Human - Reference GRCh37, Human - Celera Genomics, Human - Craig Venter (HuRef), Human - Seong-Jin Kim (Korean), Human - Chromosome 7 version 2 (TCAG), Human - unique from Reference in James Watson, YanHuang (YH; Asian), Yoruba (NA18507; African), *Mus musculus* C57BL/6J build 37, Danio rerio Tuebingen version Zv7, 18S sequences [57,317 unique sequences, 08/11/10], viral genomes in RefSeq 45 [3,761 unique sequences, 02/12/11], bacterial genomes [2,206 unique genomes, 02/12/11], archaeal genomes [155 unique genomes, 02/12/11], *Salmonella enterica* genomes [52 strains, 12/16/10] and bacterial genomes HMP [76,337 WGS sequences, 02/12/11] databases.

Assembly tools

Several assembly programs were used and compared (Figure 1). 454 reads were assembled using three different assemblers: i) Newbler, ii) MIRA with the skim_each_pass assembly option, keep_percent_memory_free at 100 and merge_xmltraceinfo (mxti)=no loadreads option (Chevreux et al. 2004), and iii) CAP3 with default settings (Huang and Madan 1999). Illumina reads were assembled using Trinity in all strategies with k-mer_method jellyfish options with 40 GB of system memory, on 4 threads (Marçais and Kingsford 2011; Grabherr et al. 2011). The longest run for the largest dataset took two weeks. Finally, MIRA and CAP3 were also used to assemble a merged dataset: raw 454 reads, assembled 454 contigs and Illumina contigs. Assemblies performed on raw data took into account quality information.

Workflow

In a first step, 454 and Illumina data were processed separately (assemblies 1 to 6; Figure 1). 454 reads were first assembled with Newbler (assembly 1), MIRA (assembly 2) and CAP3 (assembly 3). Contigs and singletons from assemblies 1 and 2 or from all three assemblies (1, 2 and 3) were pooled and reassembled with CAP3 (assemblies 4 and 5

respectively). Illumina reads were assembled using Trinity (assembly 6). In a second step, the Illumina data were merged with the 454 data to increase contig size (assemblies 7 to 10; Figure 1). In the first trial, 454 raw reads were pooled with Illumina contigs from Trinity and used as input to MIRA (assembly 7) or CAP3 (assembly 8). In a second trial, built 454 contigs from assembly 4 were pooled with Illumina contigs from Trinity and reassembled with a second CAP3 (assemblies 9 and 10). This last step led to the building of “super-contigs” which were selected when their length exceeded 200 bp using the `-min_len` option of Prinseq. Furthermore, some contigs generated from the first steps of 454 and Illumina assemblies were not reassembled in the final CAP3 run and were called contig-singlets. Some of them showed significant length and were therefore conserved (with a threshold of 700 bp). In these two last assemblies (9 and 10), potential duplicates were eliminated with CD-HIT-EST (Li and Godzik 2006).

Figure 1

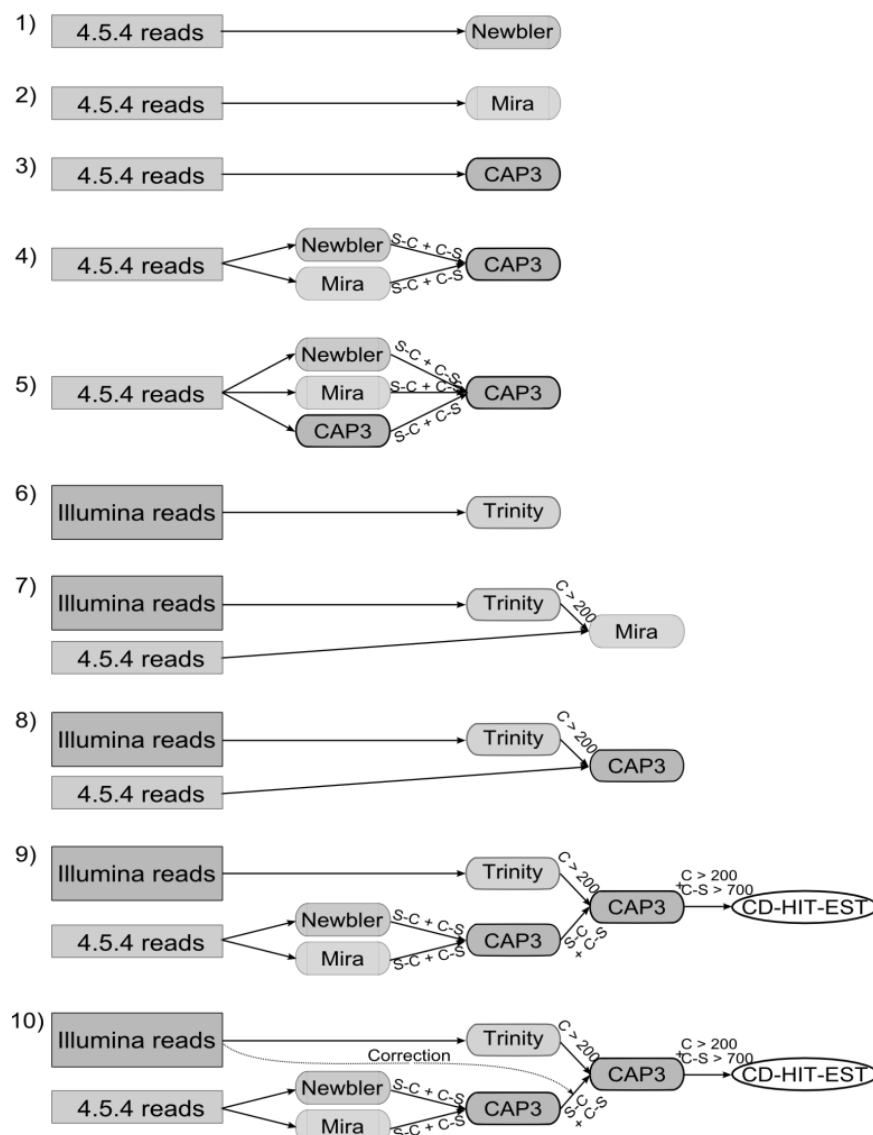


Figure 1: Workflow diagram of the different *de novo* assemblies of the *Triatoma brasiliensis* transcriptome.

“C” notifies contigs, “S-C” means super-contigs and “C-S” contig-singlets.

1) 454 reads assembled with Newbler

2) 454 reads assembled with MIRA

3) 454 reads assembled with CAP3

4) 454 contigs and singletons from assemblies 1 and 2 reassembled with CAP3

5) 454 contigs and singletons from assemblies 1, 2 and 3 reassembled with CAP3

6) Illumina reads from a pool of all populations with cleaned-1 (6a) or cleaned-2 (6b) single-end and Illumina paired-end reads of the SFF sample are assembled with Trinity (contigs higher than 200 bp are conserved)

7) and 8) After processing for assembly 6 (a/b/c), 454 read data are assembled with Illumina contigs using MIRA (assembly 7-a/b/c) or CAP3 (assembly 8-a/b/c). Contigs higher than 200 bp are selected.

9) Illumina contigs from Trinity assembly 6 (a/b/c) are merged with 454 contigs and singletons from assembly 4 and reassembled with CAP3. Contigs higher than 200 bp and singletons higher than 700 bp are selected. A CD-HIT-EST is processed to delete potential duplicates

10) Same protocol as assembly 9, but here 454 contigs and singletons from assembly 4 are corrected with Illumina reads before reassembly (see material and methods section)

Illumina data compilation for assembly

Illumina data were compiled in various ways depending on biological samples used (single individual, single population or pooled populations), sequencing approach (single-end or paired-end) or cleaning procedure. Four main ways were used for subsequent assembly: i) cleaned-1 Illumina single-end read sequences for each population (assemblies 6-a1 to 6-a14, Table 1); ii) cleaned-1 Illumina paired-end read sequences for a single individual (SFF) (assemblies 6c to 10c, Tables 1 and 2); iii) cleaned-1 Illumina single-end reads from all samples pooled (assemblies 6a to 10a, Tables 1 and 2) and iv) cleaned-2 Illumina single-end reads from all samples pooled (assemblies 6b to 10b, Tables 1 and 2). Single Trinity assembly of raw Illumina data was processed for these four types of data (assemblies 6a1-6a14, 6a/b/c). More complex workflows (assemblies 7 to 10) were only performed on data types ii) iii) and iv). To distinguish between the effect of paired-end sequencing and effect of using a single individual, assembly 6 was also conducted on data source ii) using only forward reads (assembly 6c', Table 1).

Correction of 454

454 sequencing is known to generate homopolymer tracts. Consequently, the contigs from the 454 assemblies were corrected by aligning Illumina reads on the 454 contigs using BWA (with default settings; Li and Durbin 2009). Polymorphism between 454 contigs and the mapped Illumina reads was collected with the options 'faidx' and 'mpileup' of SAMtools (using default settings; Li et al. 2009). Contigs without polymorphism with mapped Illumina reads or contigs without mapped reads were not changed. 454 contigs showing polymorphism by comparison with Illumina reads were filtered with the bcftools option and corrected: substitutions, duplications or deletions were modified following Illumina read sequences. This step was repeated five times. The correction model was applied in assembly 10.

Indices of assembly quality

Several criteria were used to evaluate the quality of the assemblies. First indices were contig lengths, including N50, total number of bases and contig number. The number of contigs was compared to the 17,155 transcripts estimated from the genome of *R. prolixus*. The completeness of the transcriptome was also estimated by the presence of sequences belonging to the set of ultra-conserved eukaryotic proteins, tested using the CEGMA pipeline (Core Mapping Approach Eukaryotic Genes; Parra et al. 2007). In addition, we created a script to check the presence of chimeric contigs by reciprocal Blast from the predicted proteome of *R. prolixus*. We selected contigs and proteins that are best aligned together in both directions and with a Blast e-value smaller than 10^{-10} . Contigs aligned with two different proteins in two *loci* without significant overlap were considered as potentially chimeric (Figure 2).

Global coverage was measured for all assemblies. The RMBT (reads mapped back to transcripts) were evaluated using BWA (Li and Durbin 2009) and SAMtools (Li et al. 2009) .

Figure 2

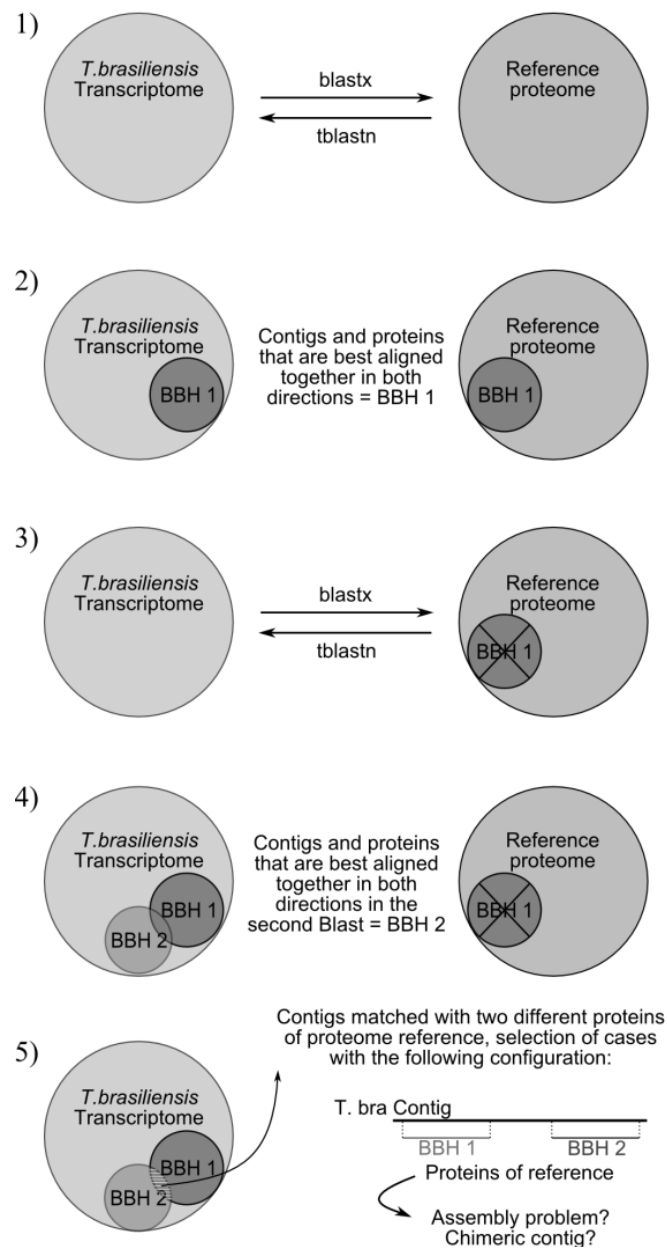


Figure 2: Research for probable chimeric contigs using reciprocal Best Blast Hit method

- 1) Reciprocal Blast between transcriptome of *T. brasiliensis* and the reference proteome of *R. prolixus*
- 2) Selection of pairs of reciprocal best hit: Contig *x* from *T. brasiliensis* and protein *y* from reference proteome formed a Best Blast Hit pair (BBH1) when protein *y* was the best hit for contig *x* in the blastx result and when contig *x* was the best hit for protein *y* in tblastn result.
- 3) Removal of proteins that were selected in step 2 (BBH1 proteins) from the reference proteome and new reciprocal blast between transcriptome of *T. brasiliensis* and partially deleted reference proteome.
- 4) Selection of new pairs of reciprocal best hit called BBH2
- 5) Selection of contigs belonging to both BBH1 and BBH2 and for which alignments of reference proteins on the contig sequence were not overlapping (less than 50 bp overlap).

General annotation

We processed and compared the functional annotation of assemblies 10a and 10c with Blast2GO (Conesa, 2005). The blast annotation part was processed in command line using Swiss-Prot database with an E-value threshold of 10^{-3} and an xml output format.

Chimeric contig validation

In order to resolve the chimeric contig problems of assembly 10-c, we studied the coverage of these contigs looking for atypical coverage profiles that could provide a clue about the misassembly location. To be sure that the atypical coverage profiles observed are specific to chimeric contigs, we compared them to the coverage of highly reliable contigs. Highly reliable contigs must have a unique alignment (with a single gene predicted from *R. prolixus* genome) of at least 700 bp over 80% of their total length and an identity rate greater than 60%. Illumina reads used for assembly 10-c were aligned with both contig categories (chimeric and reliable) with BWA. The coverage of each base was calculated using Bedtools (Quinlan and Hall, 2010) and the variation along the contig was drawn with R (R Development Core Team 2008).

SNP

To compare polymorphism level of the samples used in assemblies, we checked for SNP. For this purpose, cleaned-1 reads from i) Illumina single-end from all samples pooled and ii) Illumina single-end from a single individual (SFF) were mapped to the assembly 10c with BWA. SNP were detected from the 2 bam files generated using pileup function of SAMtools with -C 50 option to downgrade mapping quality for reads containing excessive mismatches. Quality filter of 20 was then applied on the SNP detected with VCFtools (Danecek et al., 2011).

Results

Contig number, total length and RMBT

The 454 library yielded 555,854 reads (Table 1). The fifteen Illumina libraries yielded nearly 360 million single reads and 46 million paired-end reads for nearly 41 Gb of sequences. Analysis with Deconseq revealed that probable contaminant reads varied between 0.78% and 6.85% (Table 1). Contaminant reads mainly originated from microbes that could be biologically associated with bugs. As these microbes could also be involved in the adaptation process, we maintained such reads in dataset used for assemblies.

The number of contigs varied according to assemblies (Table 2). Assemblies built only with Illumina data (assemblies 6-a, 6-b and 6-c) comprised hundred thousand contigs (256,799; 136,755 and 365,717 contigs, respectively) whereas assemblies built only with 454 reads (assemblies 1 to 5) comprised 2 to 10 times less contigs (between 15,825 and 69,917 contigs).

Among the 454 assemblies, Newbler assembly (assembly 1) had the lowest number of contigs (15,825) compared with the other 454 assemblies (58,271 to 69,917 contigs). Combining raw 454 data and assembled Illumina contigs (assemblies 7-a/b/c and 8-a/b/c) led to an intermediate number of contigs (between 56,851 and 92,040). Finally, assemblies merging Illumina and 454 contig data (9-a/b/c and 10-a/b/c) were composed of only tens of thousands of contigs (24,901 to 57,657). Preprocessing of Illumina reads also influenced the final number of contigs: the more stringent process (cleaned-2) generated less contigs (b assemblies) than the less stringent process (cleaned-1, a assemblies), whatever the assembly protocol. The total number of bases was similar between assemblies except for those obtained by Newbler (assembly 1) and Trinity (assembly 6), which were the two extremes, with 9.8 million and 179 million bases, respectively.

Among the Trinity assemblies generated from a single population data (assemblies 6-a1 to 6-a14; Table 1), assembly 6-a11 (SFC1) presented much higher number of contigs (66,830) and total number of bases (about 23 million bases) than other population samples. Inversely, assembly 6-a6 (PMD) comprised a low number of contigs (3544) and had a small total length (about 1 million of bases) probably due to the small amount of RNA extracted (Table 1). Technical replicates showed contrasted results (Table 1), the highest discrepancy being observed between SFC1 and SFC2 assemblies (66,830 contigs and 22,901,564 bp ; 10,682 contigs and 3,041,341 bp, respectively). Generally, the highest number of post-cleaned reads led to the longest assembly and may explain this discrepancy although the reverse was observed for DMB1 and DMB2. Furthermore, low sequencing depth for single population samples may explain the lack of convergence between replicates. The transcriptome from the pool of Illumina single reads for all populations presents a high number of contigs with very short lengths. A stricter cleaning of the sequences did not necessarily improve the result (assembly 6-a *versus* 6-b, Table 2). Overall, the total number of bases was directly related to the total number of cleaned reads assembled. By contrast, the Illumina paired sample SFF assembled with Trinity presented more contigs than the assembly generated from pooled population samples (365,717 compared to 256,799; Table 1), despite a smaller number of reads (15,762,629 R1 + 15,762,629 R2 *versus* 74,386,931). This also held true when SFF was assembled as single-reads. Improvement of assembly length observed from SFF data may result from i) the lower heterozygosity of the sample as a single individual was sequenced, ii) the type of tissue sequenced, as brain was added to antennae and rostrum for RNA extraction and also iii) from paired-end technology.

The RMBT is indicative of the part of the raw data that was used to build the assembly and therefore partly follows the total number of bases. For example, assembly 1 which showed the smallest number of bases assembled also showed the smallest RMBT (18.09% ; Table 2). On the other hand, assembly 6c' had the highest RMBT (71.09%) but Assembly 6c with the highest total number of bases showed a moderate RMBT (53.65%). This resulted from the highest constraint in mapping paired-end read as mapping of associated reads had to be

concordant. Except for these 3 assemblies, RBMT roughly ranged between 30% and 40% including assembly 10c (39.6%), which showed the best RBMT among assemblies combining 454 and Illumina data (assemblies 7 to 10), despite the constraint mentioned above in mapping paired-end reads.

N50 and length distribution

Length distribution of contigs from assemblies 1, 2, 4, 6-a/c, 8-a/c, 10-a/c is shown in Figure 3 and metrics for all assemblies are given in Tables 1 and 2. N50 and length distribution differed between data sources (454; Illumina single or Illumina paired) and assembly processes. Assemblies 2 to 5 of the 454 reads presented similar distributions of contig length (Figure 3) with a N50 around 500 bp (Table 2). N50 was higher for assembly 1 (662 bp) than other 454-based assemblies. In assemblies of Illumina single-end reads 6-a and 6-b, the most frequent class was between 200 and 300 bp, in agreement with N50 results (313 bp; Table 2). Assemblies of Illumina single-end reads from each population (assemblies 6-a1 to 6-a14; Table 1) or from the pool of all populations (6-a and 6-b, Table 2) presented similar N50 (about 300 bp). Assemblies performed with pooled Illumina contigs from single-end reads and 454 reads (7-a/b and 8-a/b) presented N50 and length distribution similar to 454 read assemblies alone (1 to 5). The protocol merging Illumina contigs from single-end reads and 454 contigs (9-a/b and 10-a/b) produced longer contigs (N50 from 917 to 919). Contig length distribution of these last assemblies appeared bimodal (Figure 3) due to the addition of contig-singlets longer than 700 bp (Figure 1). When similar workflows are compared, N50 were higher for assemblies generated from one individual (SFF sample) than for those from a pool of populations (6-c to 10-c *versus* 6-a to 10-a and 6-b to 10-b Table 2; 6-c and 6-c' *versus* 6-a and 6-b, Table 1).

Assembly	Contig number	Total number of bases	RMBT (%)	Length distribution (bp)				% CEGMA complete	% CEGMA partial	Potential chimeric contig proportion (%)
				max	median	average	N50			
1	15,825	9,449,447	18.09	5,141	553	597	662	26.21	37.1	0.2
2	62,028	30,302,014	33.98	4,970	410	465	491	26.61	43.55	0.05
3	58,271	25,945,159	46.30	3,936	399	445	460	25	40.73	0.04
4	63,144	31,185,087	33.63	6,234	427	493	529	28.63	44.76	0.05
5	69,917	34,003,342	34.65	6,233	432	500	535	28.63	45.16	0.05
6a	256,799	85,972,145	40.82	26,078	261	334	313	60.08	83.47	0.07
6b	136,755	45,898,586	35.08	26,182	258	335	313	43.15	71.37	0.07
6c	365,717	179,169,855	53.65	47,145	301	490	548	81.45	93.95	0.08
6c'	298,132	153,921,537	71.09	26,026	305	516	628	79.44	89.52	0.1
7a	86,985	50,616,807	35.48	20,288	511	580	628	32.66	64.11	0.09
7b	74,095	37,360,241	30.00	19,170	424	503	531	29.44	59.68	0.04
7c	92,040	60,120,211	30.17	23,712	467	653	724	64.92	75.81	0.45
8a	58,271	25,945,159	24.72	3,936	399	445	460	25	40.73	0.16
8b	56,851	28,987,179	31.14	26,188	413	509	513	47.98	67.34	0.41
8c	49,623	48,464,636	32.35	29,778	590	977	1,372	68.15	74.19	0.93
9a	33,832	28,193,255	35.42	26,085	747	833	919	66.53	72.58	0.39
9b	24,901	20,916,702	32.04	26,194	763	839	918	48.79	69.76	0.41
9c	42,752	52,356,314	35.38	47,145	923	1,225	1,393	73.39	84.27	1.05
10a	33,868	28,211,045	35.44	26,084	746	832	917	58.87	78.63	0.39
10b	24,879	20,908,511	32.09	26,193	764	840	919	48.79	70.16	0.40
10c	57,657	64,218,622	39.60	47,144	857	1,114	1,235	79.44	89.52	0.78

Table 2: Characteristics of the different assemblies

For each assembly, see Figure 1 for details of the processing

Figure 3

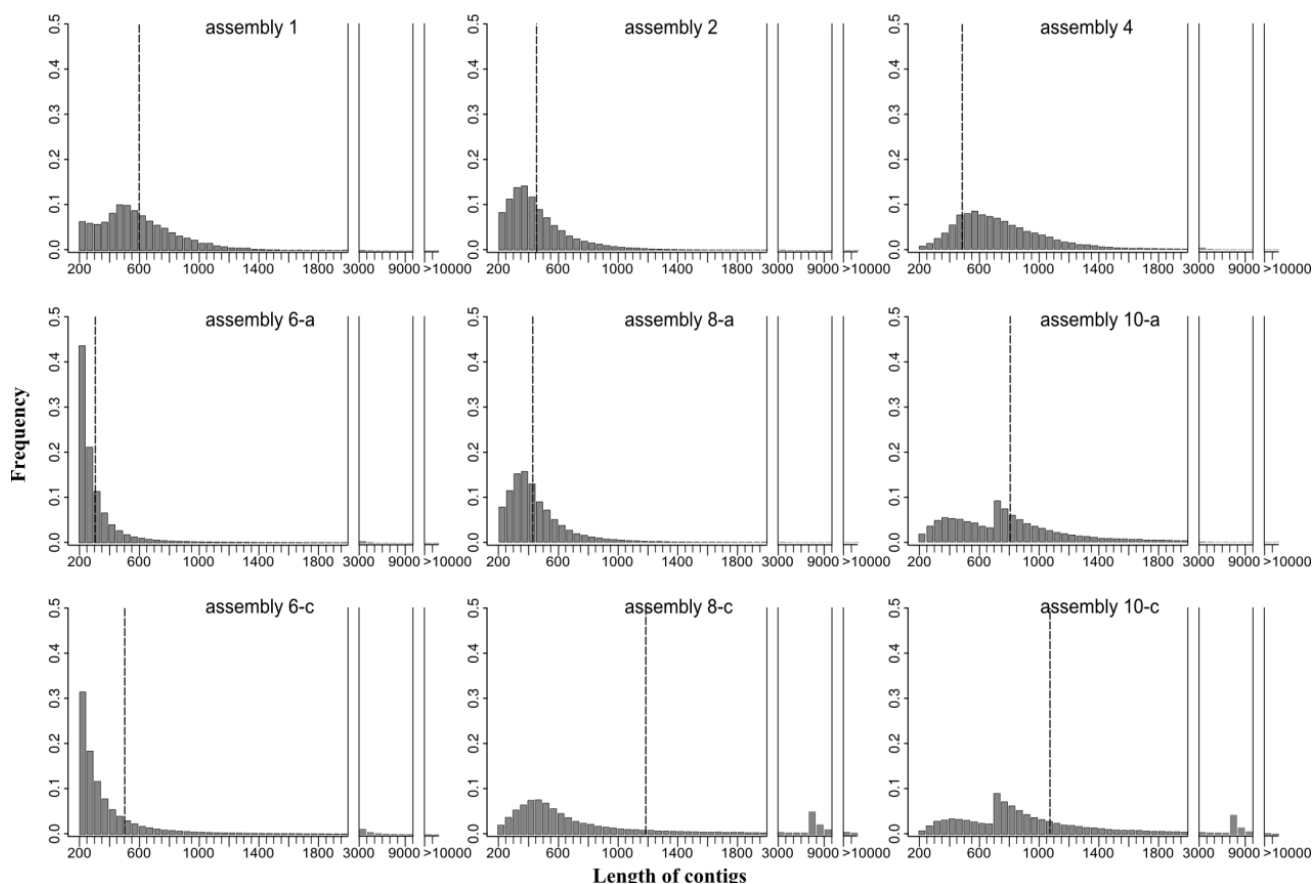


Figure 3: Length distribution of contigs from a subset of assemblies. N50 is indicated by the dotted line.

Annotation and completeness

A sequence identified as coding for one of the 248 proteins of the set of ultra-conserved eukaryotic proteins and with alignment length higher than a pre-computed minimum alignment score was named “partial”. If the alignment length was higher than 70% of the protein length it is called “complete”. The set of “partial” matches necessarily includes the set of “complete” matches. Sample SFC1 differed from other population samples in the CEGMA result (assembly 6-a11; Table 1). It presented 56% of the protein set whereas in other samples only 1.21% to 24% of the protein set were counted. The completeness was mainly related to the total number of bases in the assembly and therefore highly variable from one sample to the other even between replicates. Sample SFF from a single individual presented a CEGMA that was clearly higher (89.52%) than other samples from the pool of individuals from the same population. The best result was found for assembly 6-c: 93.95% of the set of eukaryotic conserved proteins were found and 81.45% of sequences were complete (Table 2). This improvement may result from the better quality of the assembly (low heterozygosity and paired-end data) but also from the biological representativeness of the sample: in the SFF sample, the whole head was used for RNA extraction instead of only the antennae and rostrum. The other assemblies presented from 37.1 to 89.52% of the CEGMA set and 25 to 79.44% sequences were complete (Table 2). Except for assembly 8-b, Illumina single-end reads gave better results when cleaned with the less stringent protocol cleaned-1 than with cleaned-2 (Table 2). The correction of 454 contigs with Illumina reads may increase the partial CEGMA result up to 6% (assembly 10-a *versus* 9-a, Table 2).

Number of potentially chimeric contigs

Whatever the reads (454 or Illumina single or paired), the proportion of potentially chimeric contigs remained below 0.2% (assemblies 1 to 6). By contrast, the assemblers influenced this proportion. The Newbler assembler led to a higher proportion of potentially chimeric contigs than the other 454 assemblers, CAP3 or MIRA (0.19% compared to an average of 0.05%). More significantly, the complex assemblies increased the proportion of potentially chimeric contigs up to 1% (7-c; 8-b; 9-a, b, c and 10-a, b, c).

Chimeric contigs coverage

All contigs analyzed showed heterogeneous coverage (Figure 4). Among the potentially chimeric contigs, we found three types of graphs. An example of the first type is shown with Contig 9883, where a marked drop in the average coverage is observed (around the 800th base). This region is located between the two alignments with two different *R. prolixus* putative genes. Comp1346_c0_seq1 is an example of the second type of graph: two areas are well differentiated in their coverage. The first of these areas is from the 1st to 1500th base pair and has a relatively low coverage (up to 100 reads / base) compared to the second, which extends from the 1500th to the 2500th base pair and has a higher coverage of up to 600 reads / base. These two regions correspond to those of the two Blast hits. The third type of graph is

illustrated with Contig 4175, which does not show a real difference in coverage between the two aligned regions nor a drop located between Blast hits. This last type of graph is the most frequent and is similar to the one obtained from the highly reliable contig comp1694_c0_seq1.

Figure 4

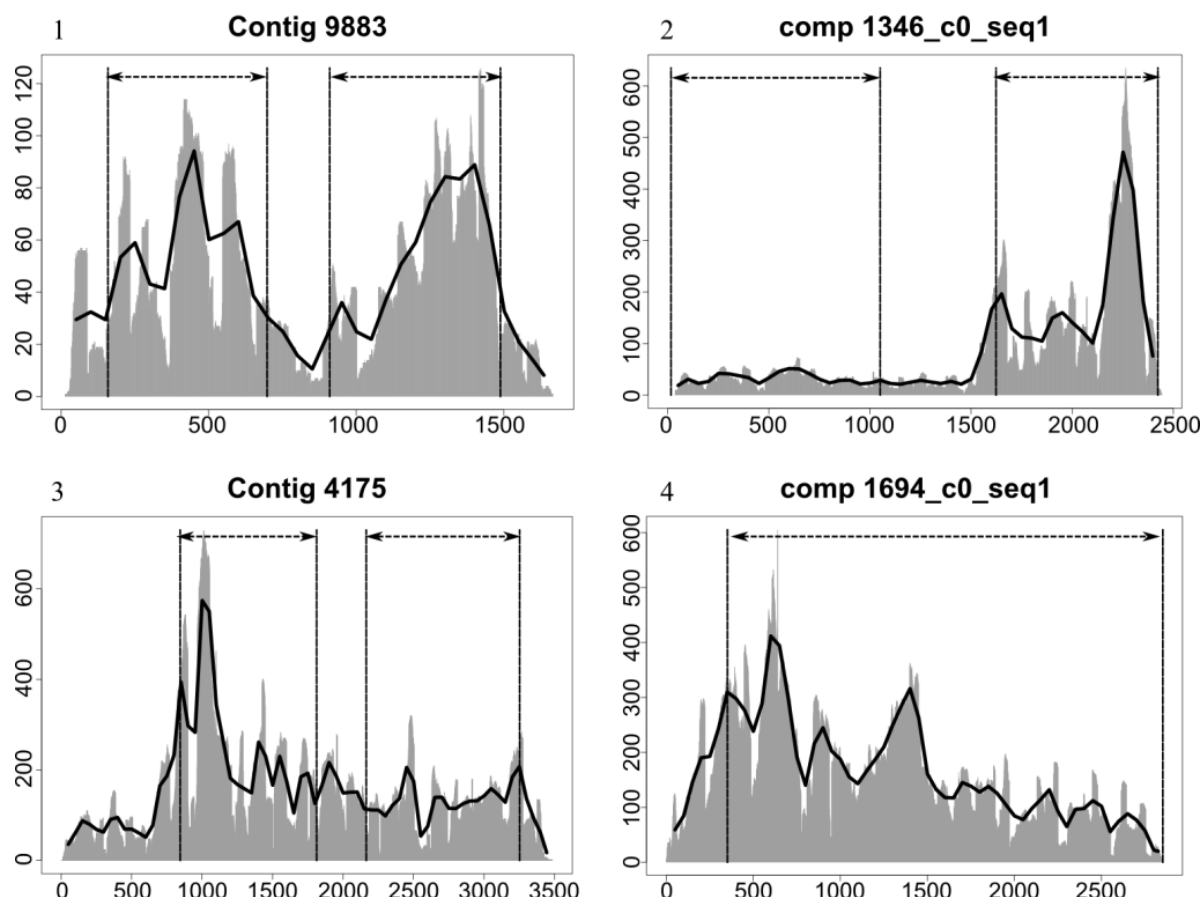


Figure 4: Graphical coverage representation. 1 to 3: coverage of 3 potential chimeric contigs, 4: highly reliable contig. The regions with a significant reciprocal Blast hit to the *R. prolixus* reference genome are bounded by vertical lines and indicated by arrows. Coverage of bases is shown in grey. The black line represents the average coverage over 100 base pair windows with a 50 bp overlap.

SNP

The population diversity was evaluated estimating the SNP number of all samples pooled and it was compared to the individual diversity estimated using the SFF sample. Mapping of single-end reads is not constrained in the same way as mapping of paired-end reads which could distort the comparison. We therefore used only forward reads of the SFF sample for SNP identification. SFF forward reads and pool of all population samples present quite similar coverage (46,051,230 versus 74,386,931 reads). We used a quality filter of 20 for SNP identification and found 53,991 SNPs for SFF and 66,279 for all samples pooled representing one SNP every 1189 bp and 969 bp, respectively. As expected, the number of SNP was lower for SFF (a single diploid individual) than for the pool data (mix of 63 diploid individuals).

However, the polymorphism level of the single SFF appears quite high compared to the population sample.

Characteristics of the best assembly

Combining the different metrics presented, the assembly showing the best compromise between contig distribution, completeness and percentage of chimeric contigs was 10-c. In order to estimate the number of the coding genes contained in the best assembly consisting of 57,657 contigs, we performed a Blast against *R. prolixus* peptide dataset. Among the 18,642 *T. brasiliensis* contigs with a peptide blast hit, 8,125 peptides were retrieved (43.6%). Assuming this non-redundant transcript rate as representative of the complete dataset, the estimated number of the coding genes is about 25,000 in this assembly.

GO annotations for assembly-10c are presented in Figure 5. Among the 57,657 contigs, 16,612 (28,8%) gave significant blast hit, 11,378 (19,7%) were aligned with a GO term and finally 7,117 (12,3%) were annotated. The biological process was composed of a majority of single-organism process (15%), metabolic process (14%), and cellular process (19%) transcripts. The cellular component was predominantly represented by organelle (28%) and cell (39%). Binding and catalytic activities represented the major molecular functions (48 and 36% respectively). One contig was annotated as virion cellular component. Similar results were found for assembly 10a (Supplementary data S2).

Figure 5

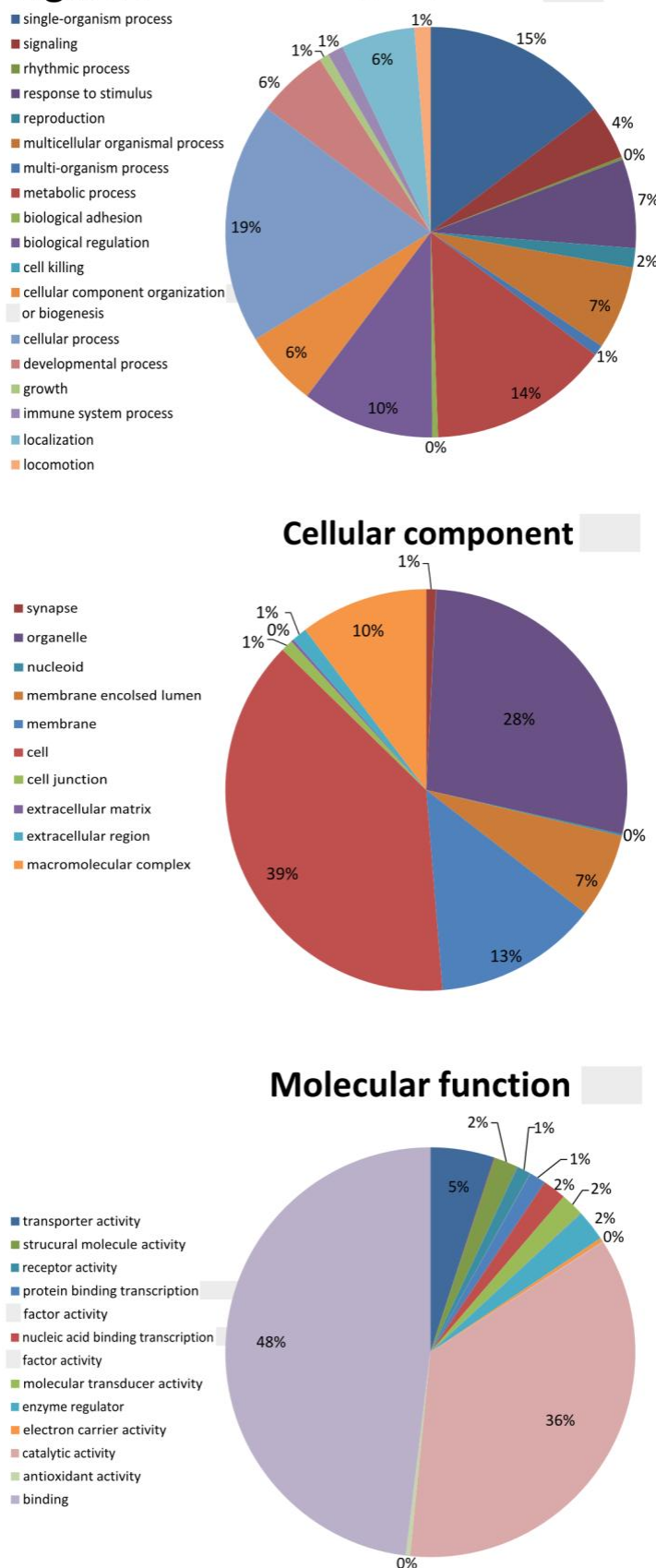


Figure 5: Distribution of *T. brasiliensis* contigs from assembly 10c annotated at GO level 2.

Discussion

In a comparative study of non-model organism genomics, the conclusions drawn from the analysis of HTS data partly depend on the quality of the assembly (Stapley et al. 2010; Vijay et al. 2013). Major information may be missed out from an incomplete assembly. Moreover, transcripts that are improperly assembled and/or scattered across different contigs may complicate the read mapping and skew differential expression results. Thus, it is fundamentally important to construct the most optimized assembly from an available dataset. In order to conduct a comparative study of the different *T. brasiliensis* populations in the future, the first step was to assemble the transcriptome of this species. In this article, we tested and compared several assembly methods from different datasets. The quality of these assemblies was assessed according to numerous criteria: number of transcripts, transcript size, completeness, and chimeric contig occurrence based on the predicted proteome of *R. prolixus*.

Influence of assembly and workflow software

Concerning 454 data assemblies, Newbler generated longer contigs than MIRA (N50 of 662 bp *versus* 491 bp). Cahais et al. (2012) found similar results when assembling 454 reads from *Ciona intestinalis*, as did Ren et al. (2012) on simulated data and 454 reads from *Trychophyton rubrum*. Newbler is more stringent in filtering contig size. The size distribution appeared therefore moved toward longer contig size (increased N50) but this resulted mainly from the elimination of small contigs as well as the genetic information they included (Cahais et al. 2012). This is typically in line with what we observed in assembly 1. Our results suggest that, despite being longer, Newbler contigs were less informative than those generated using MIRA (CEGMA partial 37 *versus* 43.55 %), in agreement with the results of Ren et al. (2012) and Cahais et al. (2012). However, in this latter study, the highest N50 and contig number aligned to the reference were found using CAP3. In our study, CAP3 produced similar results to those of MIRA. Several software programs are also available to assemble Illumina data. According to Zhao et al. (2011), the assembler Trans-Abyss is a good compromise between RAM need, runtime and assembly quality. Cahais et al. (2012) also endorse this assembly software. Oases is another assembler frequently used (Feldmeyer et al., 2011; Francis et al., 2013; Haznedaroglu et al., 2012; Zhao et al., 2011). We tested Oases and Trans-Abyss to assemble *T. brasiliensis* Illumina reads with different k-mer sizes and default options. The assemblies generated (data not shown) were either similar or less informative than those obtained with Trinity. Therefore, we decided to use only this latter software.

Completeness and annotation of the transcriptome

The expected distribution of transcript number and length may be estimated from the reference genome of *R. prolixus*, which has been recently sequenced and assembled (www.vectorbase.org). The predicted transcriptome of this species comprises 17,155 genes with a mean transcript length of 1016.5 bp and a maximum length of 34,028 bp. Therefore, we expected to find the same orders of magnitude for *T. brasiliensis* transcripts. All transcriptome assemblies (except

assembly 1) had a contig number and a total number of bases higher than that of *R. prolixus* predicted gene set defined from the genome. Transcriptome assemblies from other Hemiptera species also exhibit a large number of contigs. For example, assembly of transcriptome data comprised 54,748 contigs for *Trialeurodes vaporariorum* with an average length of 965 bp (Karatolos et al., 2011) and 35,646 expressed sequence tags with an average contig length of 759 bp in the bed bug *Cimex lectularius* (Bai et al., 2011). In more distant arthropod groups, 51,999 contigs were assembled for *Sesamia nonagrioides* (Glaser et al., 2013), 128,391 for *Theridion grallator* and 104,481 for *Theridion californicum* (Croucher et al., 2013). The discrepancies between transcript and gene numbers could be attributed to alternative splicing events producing several different transcripts from a single gene and increasing the transcriptome size. Such a result was observed by Hull et al., (2013) who found 36,131 transcript isoforms corresponding to 19,742 genes in western tarnished plant bug *Lygus hesperus*. However, the estimated gene number of *T. brasiliensis* transcriptome (about 25,000) was still higher than the one of *R. prolixus*. The gene content between the two bug species may be somewhat different as they are quite distant: genetic duplications specific to *T. brasiliensis* may have occurred followed by gain of function. The difference may also result from an underestimation of the gene content of the recently sequenced genome of *R. prolixus*. A CEGMA test we performed on the predicted transcriptome of *R. prolixus* gave a result of 14.11% complete proteins and 17.34% partial proteins. This suggests that the putative gene set of *R. prolixus* is still incomplete.

According to CEGMA, assemblies from 454 (assemblies 1 to 5, Table 2) and those from Illumina single-reads of populations (assemblies 6-a1 to 6-a14, Table 1) gave very poorly informative results. Overall, the CEGMA increased with the number of initial sample reads. The transcriptome built from pool of single reads from all population samples (assembly 6-a) has a high completeness rate (CEGMA: 83.47% partial and 60.08% complete), but one lower than the assembly constructed from a single sample (SFF), while the sequencing depth of this latter sample is lower. According to Francis et al., (2013) the transcriptome completeness using CEGMA values increases with the sequencing depth, but decreases beyond a threshold (estimated between 20 and 30 million reads). An excess of reads may complicate assembly and lead to errors due to the occurrence of misassembled or chimeric cases. However, transcriptomes of spiders *Theridion grallor* and *Theridion californicum* were assembled from about 250 million Illumina reads each and showed CEGMA results of 99% and 98% complete, respectively (Croucher et al. 2013). It is more likely that appropriate read number to maximize the assembly completeness is variable and intrinsic to each organism depending on the complexity of the genome, such as the number of transcribed genes, their average size and the gene expression level (Zhao et al. 2011). Moreover, the completeness of the transcriptome also varies depending on the tissues targeted for the extraction. Francis et al. (2013) found a higher percentage of conserved eukaryotic orthologous genes using a transcriptome built from a complete individual RNA than from a single tissue (90% versus 63–81 %), reflecting that genes

may not always be expressed at a sufficient level in all tissues. Our population samples were derived from antennae and mouthparts of bugs, while brain tissue was also targeted for the SFF sample in order to obtain enough RNA material from a single individual. This addition of brain tissue may have improved the completeness of the transcriptome.

For this SFF sample, two other parameters could also explain its CEGMA value. First, this sample comprises a single individual, which decreases the polymorphism rate and facilitates assembly. Indeed, SNP identification reveals a higher polymorphism in the pooled population sample than in the individual sample. Second, the paired-end technology used for this assembly has a positive effect on the completeness of the transcriptome as revealed by the comparison with the single-end assembly for this sample. According to (Sengupta et al., 2011), paired-end technology makes it possible to detect alternative splice junctions, insertions and deletions, and facilitates the *de novo* assembly of the transcriptome. However, single-end sequencing is still very widely used because of its lower cost.

Despite differences in completeness and polymorphism, the type of samples did not change the GO annotation. The proportion of biological process, cellular composition and molecular function annotations are similar between assemblies 10-a and 10-c and comparable to those found for other chemosensory transcriptomes of insects, such as *Sodoptera littoralis* (Poivet et al., 2013) and *S. nonagrioides* (Glaser et al. 2013). Equivalent repartitions were also found in transcriptomes from guts or full-body of *Aphis glycines* (Liu et al., 2012b).

Chimeric contigs

Among the assemblies tested, the characteristics of assembly 10-c appeared to offer the best compromise, taking into account the contig numbers, the N50 and the CEGMA values. This assembly generated contigs up to 47,145 bp, a number compatible with the largest predicted transcript for *R. prolixus* (34,028 bp, www.vectorbase.org), but the occurrence of chimeric contigs could not be excluded. Overall, the number of potential chimeric contigs increased with the complexity of the assembly workflows (assemblies 7 to 10), but also in relation to the number of available annotated genes. For example, assemblies of 454 EST simulated from the human genome generated with Newbler, CAP3 or MIRA gave high percentages of chimeric contigs (14% with MIRA, 15% with CAP3 and 38% with Newbler; Mundry et al. 2012). However, in the Mundry et al. (2012) study, assemblies and reference are issued from the same genome, which enhances the power of chimera detection. In our case, the reference genome is quite distant, which should reduce the percentage of chimeric contigs detected. Conversely, the assembly and annotation of the *R. prolixus* genome is unfinished, which could artificially increase the percentage of false chimeric contigs. Owing to both observations, the number of potential chimera detected (up to 1%) seems satisfactory.

The occurrence of chimeric contigs had been examined by studying their coverage. For some of the chimeric contigs, the significant drop in coverage could be interpreted as the boundary between two different transcripts assembled in a single contig. For some others, two regions

were differentiated by distinct coverage that may be a consequence of two differentially expressed transcripts being erroneously assembled. It could be also the result of alternative splicing with the presence of an exon over-expressed compared to others because it is included in several isoforms. Even in this case where the chimera nature of the contig is not established, it may be preferable to separate both exons to better estimate differential expression level. That is why, in these two coverage profiles, chimeric contigs were cut in order to generate two unlinked transcripts. However, in most cases, the coverage of potentially chimeric contigs was very heterogeneous with no clear boundary. Such heterogeneity was also found in contigs aligned on only one reference gene with a high reliability, which questions the chimeric nature of this contig class. Moreover, in some rare cases, the fusion of mRNA can be generated during trans-splicing events (Bonen, 1993). Such fusion would be interpreted as chimeric contigs in our protocol while it corresponds to a true transcript. Consequently, the percentage of potential chimeric contigs is probably overestimated.

Conclusion

Our study documents how a high quality assembly results from various parameters including the sampling procedure, the sequencing strategy and the assembly workflow. Despite it is tempting to perform assembly on population pool with the aim to better appreciate global genetic information, we do not recommend such an approach. Instead, we showed that the use of a single individual with large representation of biological tissues is preferable to improve both the total number of bases assembled and the completeness of the assembly. In this case, the assembly procedure is facilitated by decreasing polymorphism rate and the representation of the whole transcriptome is improved by the diversity of tissues sampled. Sequencing technologies now require small amount of RNA allowing to sequence only one individual even for species or tissues of small size. In addition to the choice of biological material, we highly recommend a mixed sequencing strategy merging both long reads as generated by 454 or the recent MiSeq Illumina technology (that allows up to 2x 250pb fragments in paired-end reads) and short but paired-end Illumina reads (2x 100pb fragments). We showed that the well-documented biases of 454 technology (such as homopolymers) may be corrected using Illumina sequences and that utilization of both technologies greatly improves length distribution of contigs generated. The size distribution is also increased by the usage of paired-end rather than single-end technology. Finally, the usage of various assemblers is also a powerful approach combining the specific advantages of each as far as duplicate contigs are eliminated in the final step.

Acknowledgments

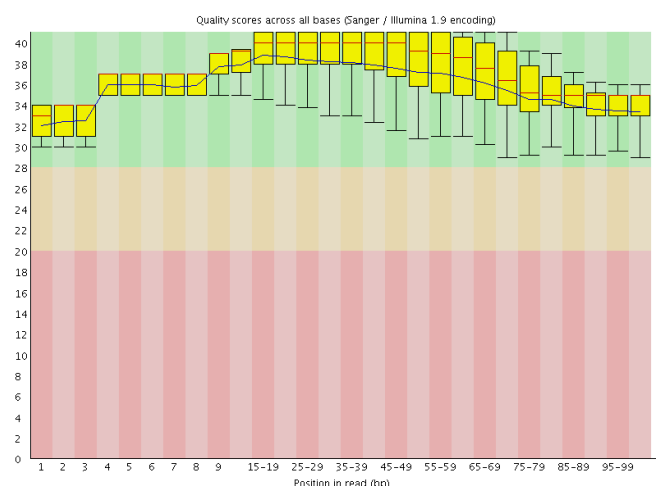
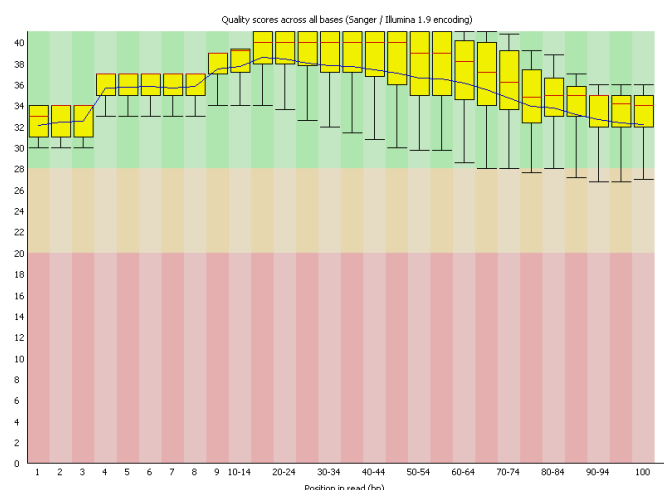
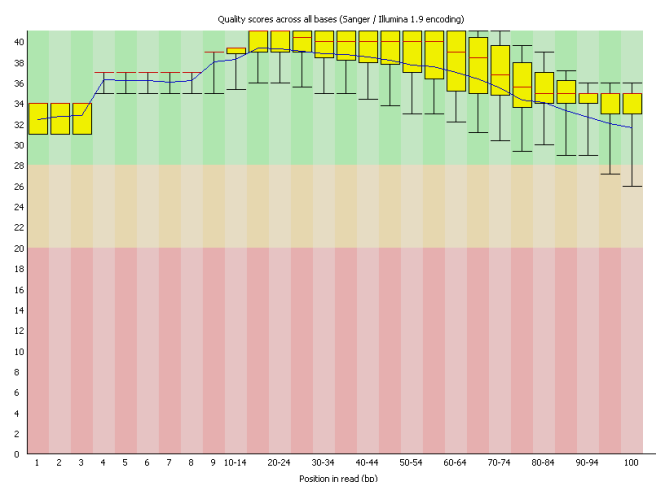
We would like to thank Rachel Legendre and Claire Toffano of Institut de Génétique et Microbiologie CNRS - UMR 8621 who gave us the script for 454 contig correction. We thank Marie-Christine François (iEES, INRA Versailles, France) for help with the *T. brasiliensis* RNA extractions. The authors are also very grateful to the engineers of the bioinformatics platforms Genouest at the University of Rennes 1 and eBio of the University Paris Sud for technical support. This work has benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr). This study was funded by the French Agence Nationale de la Recherche (ADAPTANTHROP project, ANR-097-PEXT-009) and supported by the labex Biodiversité, Agroécosystèmes, Société, Climat (BASC; University Paris Saclay, France). Marchant A. was funded by the Idex Paris Saclay, France.

Conflict of interest

The authors announce that they have not a financial relationship with the organization that sponsored the research.

The authors declare that they have no conflict of interest.

Supplementary material

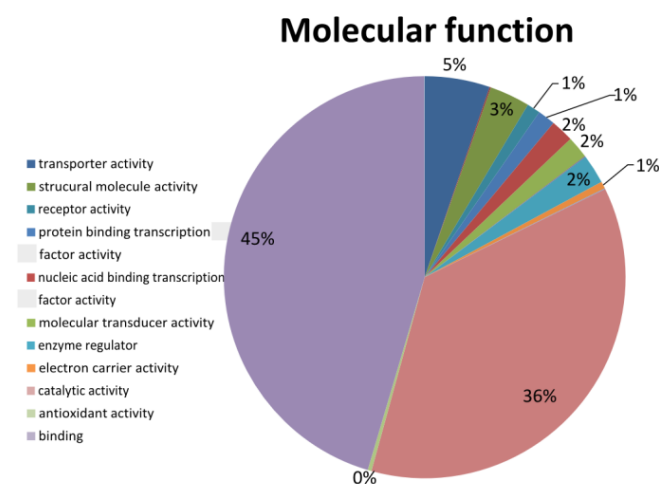
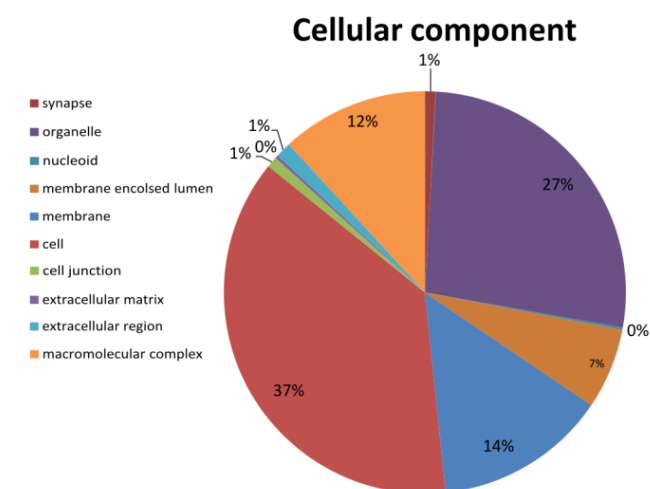
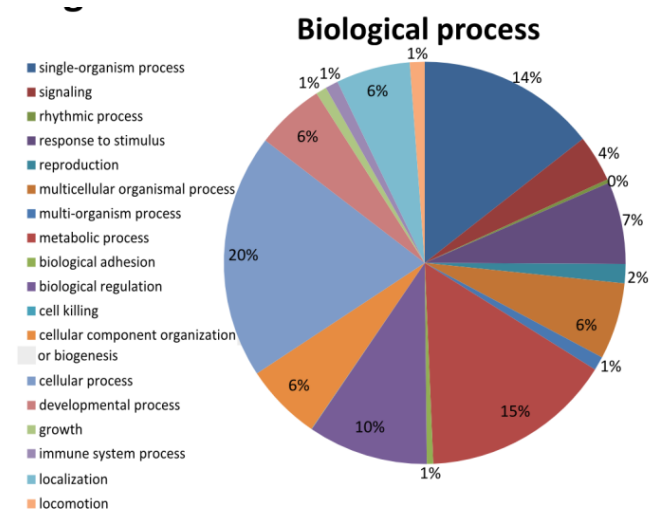


S1: FastQC report of DMB1 sample.

A: Before preprocessing (adapter clipped only);
29,820,509 sequences

B: After preprocessing cleaned-1; 4,197,000 sequences

C: After preprocessing cleaned-2; 3,311,104 sequences



S2: Distribution of *T. brasiliensis* contigs from assembly 10a annotated at GO level 2.

Chapitre 2 : Analyse de l'expression différentielle entre populations de *T. brasiliensis* provenant d'environnements différents

Résumé en français

Introduction : Des programmes de contrôle par éradication chimique des vecteurs de la maladie de Chagas ont été mis en place et ont permis une importante diminution du nombre de personnes infectées. Cependant, la domiciliation, c'est-à-dire la colonisation de l'habitat humain, est observée pour les nouvelles espèces (*Triatoma brasiliensis* en est un exemple). Il est donc impératif de comprendre ce processus afin d'améliorer les méthodes de lutte contre les vecteurs. Le système chimiosensoriel qui permet aux insectes d'interagir avec leur environnement a un rôle essentiel dans l'adaptation des insectes. Aussi, nous avons mené une approche transcriptomique descriptive et comparative sur le transcriptome chimiosensoriel de plusieurs populations de *T. brasiliensis* provenant de différents écotypes (sylvatique, péri-domicilié et domicilié). Cette analyse a pu être réalisée suite à l'élaboration du transcriptome de référence de *T. brasiliensis* décrite au chapitre précédent.

Matériels et Méthodes : La stratégie d'assemblage sélectionnée à partir de l'étude précédente a été conservée. Cependant, afin de faciliter les comparaisons des données transcriptomiques, seul le plus grand isoforme a été gardé avant la mise en commun des données Illumina avec le 454. L'annotation des contigs codant pour des protéines de liaison (OBPs et CSPs) a été réalisée par Blast et validée par recherche des motifs conservés sur les séquences traduites en acides aminés (présence du signal peptidique, d'hélices α des cystéines caractéristiques de ces protéines). Après traduction de ces contigs en protéines, ces séquences protéiques ont par la suite été alignées avec celles d'autres Paraneoptera pour réaliser un arbre phylogénétique. Des analyses d'expression différentielle ont été effectuées entre plusieurs échantillons de *T. brasiliensis* en tenant compte de leur environnement (sylvatique, péri-domicilié et domicilié) et de leur sexe. Deux pipelines R ont été utilisés et comparés: DESeq2 et EdgeR. Ces analyses ont été complétées par une approche de regroupement des transcrits en fonction de leur profil d'expression à l'aide du pipeline HTSCluster.

Résultats : L'annotation a révélé 25 contigs codant des OBPs et 16 codant des CSPs. Nous avons également annoté 22 contigs codant les protéines *takeout*, 37 contigs codant des protéines impliquées dans la protection contre la pénétration de toxines ou dans la détoxification comme les protéines cuticulaires, 71 contigs codant des cytochromes P450 et 5 contigs codant des glutathione S-transférases. L'analyse phylogénétique des transcrits traduits en protéines a montré l'expansion de ces gènes chez les Triatominae par rapport aux autres Paraneoptera. De nombreux contigs (4024) ont été identifiés comme différentiellement exprimés (DE) entre mâles et femelles. Par ailleurs, les punaises hématophages domiciliées ont montré des profils d'expression différents de ceux des punaises sylvatiques (325 gènes DE) et péri-domiciliées (3577 gènes DE). Parmi les gènes DE, et généralement sous-exprimés chez les punaises domiciliées et chez les femelles, nous avons clairement identifié, des gènes appartenant

au système chimiosensoriel (principalement des OBPs et CSPs), mais également des gènes codant les protéines *takeout* impliquées dans le comportement alimentaire des adultes et le rythme circadien.

Conclusion : Les gènes chimiosensoriels OBPs et CSPs pourraient être de bons gènes candidats impliqués, par leur plasticité ou par sélection naturelle, dans l'adaptation des punaises à un nouvel environnement anthropique. L'environnement domestique est probablement moins diversifié en xénobiotiques et plus stable en ce qui concerne les paramètres abiotiques que les environnements péri-domestiques ou sylvatiques, ce qui pourrait expliquer la plus faible expression génique mesurée chez punaises domiciliées.

Article 2: Under expression of chemosensory genes in domiciliary bugs of the Chagas disease vector *Triatoma brasiliensis*

Authors

A. Marchant¹, F. Mougell¹, E. Jacquin-Joly², J. Costa³, C. E. Almeida⁴, M. Harry ^{*1}

(1) UMR Evolution, Génomes, Comportement, –Ecologie, CNRS-IRD- Univ. Paris-Sud, Université Paris Saclay, Campus CNRS, Bâtiment 13, BP1 - 91198 Gif-sur-Yvette – France

(2) INRA, UMR 1392, Institut d'Ecologie et des Sciences de l'Environnement de Paris, Route de Saint Cyr, 78026 Versailles, France

(3) Laboratório de Biodiversidade Entomológica; Instituto Oswaldo Cruz - Fiocruz; Rio de Janeiro, RJ; Brasil Instituto Oswaldo Cruz, Fiocruz – Brazil

(4) Departamento de Ciências Biológicas, Faculdade de Ciências Farmacêuticas (UNESP), Araraquara, SP; Brasil UNESP – Brazil

***Myriam Harry**

Email : myriam.harry@egce.cnrs-gif.fr

Abstract

Background

In Latin America, the bloodsucking bugs Triatominae are vectors of *Trypanosoma cruzi*, the parasite that causes Chagas disease. Chemical eradication programs have been launched to control Chagas disease vectors. However, the disease persists because, once vectors were eradicated, new vectors from natural habitats were able to recolonize houses—a process called domiciliation. *Triatoma brasiliensis* is one example. Because the chemosensory system allows insects to interact with their environment and plays a key role in insect adaption, we conducted a descriptive and comparative study of the chemosensory transcriptome of *T. brasiliensis* populations from different ecotopes.

Methodology/Principal Finding

On a reference transcriptome built using *de novo* assembly, we found 27 Odorant Binding Proteins (OBP), 17 Chemosensory Proteins (CSP), 25 takeout proteins, 72 cytochrome P450s, 5 glutathione S-transferases, and 49 cuticular protein. Using gene phylogenies, we showed that most of the OBPs and CSPs for *T. brasiliensis* had well supported orthologs in *R. prolixus*. We also showed clear expansions within the bloodsucking bugs and more generally within all Hemipterans compared to the other species in the super-order Paraneoptera.

Using both DESeq2 and EdgeR software, we performed differential expression analyses between populations of *T. brasiliensis*, taking into account their environment (sylvatic, peridomiciliary and domiciliary) and sex. We also searched clusters of co-expressed contigs using HTSCluster. Among differentially expressed (DE) contigs, most were under-expressed in the chemosensory organs of the domiciliary bugs compared to the other populations and in females compared to males. We clearly identified DE genes that play a role in the chemosensory system.

Conclusion/Significance

Chemosensory genes could be good candidates for genes that contribute to adaptation or plastic rearrangement to an anthropogenic system. The domiciliary environment probably

includes less diversity of xenobiotics and probably has more stable abiotic parameters than do sylvatic and peridomiliary environments. This could explain why both detoxification and cuticle protein genes are expressed less in domiciliary bugs. Understanding the molecular basis for how vectors adapt to human dwellings may reveal new tools to control disease vectors; for example, by disrupting chemical communication.

Author Summary

In Latin America, bloodsucking bugs are vectors of *Trypanosoma cruzi*, the parasite that causes Chagas disease, which is a major public health problem. Though chemical control campaigns have been effective against vectors, the disease persists because new vectors from natural habitats have been able to recolonize human habitations. This is true of *Triatoma brasiliensis*. Its capacity to adapt to a new habitat could be linked to changes in the repertoire of chemosensory system genes, particularly odorant binding proteins (OBP) and chemosensory proteins (CSP), which are important for detecting odor stimuli.

This study looks at the chemosensory system of Triatominae in an attempt to document the adaptation process and the domiciliation of disease vectors. We used transcriptomic data obtained using high-throughput sequencing to annotate and list the chemosensory genes. We also show differential gene expression in *T. brasiliensis* populations from different habitats.

Introduction

Chagas disease is a potentially fatal parasitic disease caused by *Trypanosoma cruzi*, an endemic kinetoplastida that has infected five million people in Latin America (WHO, 2015) and is transmitted by blood-sucking bugs (Hemiptera, Reduviidae, Triatominae). Chemical control campaigns against Chagas disease vectors have considerably reduced its prevalence in recent decades by eradicating populations of the most common vector species in human habitats. *Triatoma infestans* has been eradicated in Southern Cone countries and *Rhodnius prolixus* in parts of the Andean Pact and Central America. However, investment in vector control and surveillance has decreased in Brazil since these target insects were eradicated, and now several other triatomine species—like *Triatoma brasiliensis*—are increasing in abundance in human habitats, taking over the ecological niches left open by the eradication programs. Some of these sylvatic populations have been able to expand to colonize domiciliary and peridomiliary environments (Almeida et al., 2008; Borges et al., 2005; Costa, 1999; Costa et al., 2003, 1998). We have a critical need to understand how vectors adapt to human environments—a process called domiciliation—to control the spread of Chagas disease.

The insect chemosensory system plays a key role in ecological adaptation to new or changing hosts or habitats (see Benton, 2015). At the molecular level, chemical recognition depends on the activation of specific sets of genes, including genes encoding odorant-binding

proteins (OBPs), chemosensory proteins (CSPs), olfactory receptors (ORs) and gustatory receptors (GRs) (reviewed in Leal et al., 2013). OBPs and CSPs are soluble, secreted proteins that appear to play a role in ligand binding and transport to membrane receptors. ORs and GRs recognize specific ligands and transform the chemical signal into an electrical signal that will be transmitted to the brain, leading to the insect response (Bohbot and Pitts, 2015; Jacquin-Joly and Merlin, 2004; Sánchez-Gracia et al., 2009). Insect OBPs (Vieira et al., 2007), ORs (Guo and Kim, 2007) and GRs (Engsontia et al., 2014) evolve rapidly via gene duplication/loss events, in parallel with adaptation to new ecological niches. Changes in olfactory sensitivity can be driven by these gain and loss events but also by intragenic mutations (Leary et al., 2012) or variation in gene expression (see McBride et al., 2014).

The insect odorant landscape also includes pheromones—some crucial for mating (Karlson and Luscher, 1959), as demonstrated for *T. brasiliensis* (Vitta et al., 2009). The expression of transport proteins of these specific odorants may therefore differ between sexes (Györgyi et al., 1988; Krieger et al., 1996, 1991; Raming et al., 1989; Vogt et al., 1991; Vogt and Riddiford, 1981).

When their environment changes, organisms respond by tuning gene expression. Rapid response to a brief, stressful event can persist as a long-term adaptation to a selective pressure (López-Maury et al., 2008). A change in gene expression is a major component of genetic modulation in phenotypic evolution (Holloway et al., 2007). New generations of sequencing have considerably expanded opportunities to explore transcriptomes of non-models organisms using RNA-seq. This revolutionary tool provides unprecedented precision in the measurement of transcript levels (Wang et al., 2009).

The aim of the present study was to detect differentially expressed genes that play a role in the domiciliation process and sexual behavior of *T. brasiliensis* populations sampled in different ecotopes (sylvatic, peridomiciliary, domiciliary). We first evaluated the diversity of OBPs and CSPs through the *T. brasiliensis* reference transcriptome (Marchant et al., 2015a), and expansion/loss events in the super-order Paraneoptera by building protein phylogenetic trees. We then evaluated contigs that were significantly differentially expressed (DE) and searched for contig clusters that show similar expression patterns in different environmental conditions using HTScluster. We showed that the following genes were significantly differentially expressed across genders and ecotopes: genes belonging to the chemosensory system (especially OBP and CSP genes), genes encoding *takeout* proteins involved in adult feeding and male courtship behavior, and genes encoding for proteins involved in detoxification or in preventing toxins from penetrating the cuticle.

Materials and methods

Sampling, RNA extraction and sequencing

Four populations of *T. brasiliensis* were collected in Caicó city, Rio Grande do Norte, Brazil (between 06°23' to 06°41'S and 36°58' to 37°12'W; Table 1), within the *Caatinga* ecoregion: i) population B was sampled in domiciles (Dom), which were the indoor spaces of homes where triatomines are generally found in the crevices of mud walls, in furniture and under beds; ii) population D was sampled in peridomiciliary (Per) areas, which were the areas outside and within approximately 100 m of homes, where domesticated animals sleep or are maintained; and iii) populations A (06°28' 25''.0 S / 37°05' 21''.4 W) and C (06° 28' 21''.6 S / 37° 05' 12''.5 W) were about 1.5 km apart in sylvatic (Syl) areas in the Environmental Preserved Area (EPA) of Caicó that is under the supervision of military guards. The approximate linear distances (in km) between populations were as follows: 26 km between B and A-C, 20 km between B and D, and 7 km between D and A-C. Domiciliary and peridomiciliary samples were collected in the daytime; sylvatic samples were collected at night. Males and females were separated in all four populations, resulting in eight samples. To target the chemosensory system, we extracted RNA from the antennae and rostrum using the TRIzol® Reagent kit (Invitrogen, Carlsbad, CA, USA). We pooled 4 to 15 individuals of the same sex and from the same population to ensure there was enough RNA for sequencing and to measure average gene expression in a population. We made technical replicates for samples with enough RNA. Fourteen libraries were constructed using TruSeq RNA Kit 2010 from Illumina, and two runs of each were sequenced with the Illumina HiSeq 2000 method in single-reads of 100 bp on the LGC Genomics platform.

The number of reads obtained for each sample, replicate and run is reported in the Table 1.

Sample	Environment	Sex	No. of individuals	No. of filtered adapters reads			No. of mapped reads				
				Run1	Run2	Total	Run1	Run2	Total	Percentage reads mapped/total	Max counts (%)
DFB	D	F	4	9,210,196	13,482,702	22,692,898	1,391,136	1,985,127	3,376,263	14.88	28.41
DMB1	D	M	5	13,945,801	29,820,509	87,001,857	4,198,630	9,667,151	26,453,737	30.41	38.81
DMB2	D	M	5	24,053,273	19,182,274		6,852,782	5,735,174			
PFD1	P	F	6	7,024,781	15,993,588	45,152,589	2,728,046	6,627,049	17,319,989	38.36	13.51
PFD2	P	F	6	10,559,081	11,575,139		3,691,981	4,272,913			
PMD	P	M	11	1,851,110	4,771,764	6,622,874	550,846	1,460,386	2,011,232	30.37	34.49
SFA1	S	F	8	5,122,662	20,488,218	48,421,745	1,849,122	7,924,973	16,626,180	34.34	22.21
SFA2	S	F	8	7,782,766	15,028,099		2,229,032	4,623,053			
SMA1	S	M	10	25,231,786	3,319,544	74,845,155	9,460,295	1,285,816	27,848,905	37.21	28.87
SMA2	S	M	10	24,253,202	22,040,623		8,847,900	8,254,894			
SFC1	S	F	4	31,123,710	40,633,951	94,201,052	5,794,452	7,943,048	16,898,699	17.94	20.88
SFC2	S	F	4	3,421,756	19,021,635		494,700	2,666,499			
SMC1	S	M	15	13,953,748	21,062,918	46,892,268	4,876,329	8,285,962	19,181,873	40.91	46.26
SMC2	S	M	15	4,593,481	7,282,121		2,260,114	3,759,468			

Table 1: *T. brasiliensis* samples from which chemosensory transcriptomes were sequenced

Sample names are made up of three letters that indicate the environment (D = domiciliary, P = peridomiciliary or S = sylvatic); the gender (M = male or F = female); and the population index. The number at the end of the sample names indicate technical replicate number (no replicates were performed for DFB and PMD F). RNA was extracted from antennae and rostra, sequenced in Illumina single-reads and mapped in the reference transcriptome. Number of filtered reads and number of mapped reads per sample are indicated. The Max_counts represents the percentage of reads mapped in the most expressed contigs. S = sylvatic, P = peridomiciliary and D = domiciliary

***De novo* assembly of the reference transcriptome**

We assembled the reference transcriptome by following procedure 10c published by Marchant et al. (2014) with the following modifications: we cleaned paired-ends together using prinseq version 0.20.4 and we added the step of selecting the longest isoform per “gene” (defined as transcript cluster by Trinity) from the Trinity output. The RNA dataset consisted of one chemosensorial library of a sylvatic female sequenced in Illumina paired-end and of eight chemosensorial libraries from the males and females of the four populations described above sequenced with 454 technology (Marchant et al., 2015a). Transcriptome completeness was assessed using the CEGMA (Core Eukaryotic Genes Mapping Approach) pipeline that searches for the presence of sequences belonging to a set of ultra-conserved eukaryotic proteins (Parra et al., 2007).

Annotation

We annotated contigs that code for a set of proteins of interest (CSPs, OBPs, takeout proteins, cytochrome P450, glutathione S-transferase, cuticular proteins) in the reference transcriptome. We queried insect sequences retrieved from GenBank against our reference

transcriptome using Tblastn searches with an e-value threshold of 10^{-6} . The selected contigs were aligned to the non-redundant protein database (Blastx with an e-value threshold of 10^{-6}). Only contigs with at least one protein match from the families cited above were kept.

For the OBP and CSP protein families, contigs were translated and checked for the following: conservation of 6 (OBP) and 4 (CSP) cystein positions, the presence of α helices using PSIPRED (McGuffin et al., 2000), and the presence of a signal peptide using SignalP (Petersen et al., 2011). The OBP/CSP gene repertory of *T. brasiliensis* was compared to those of several Paraneoptera. To do this, we selected translated sequences from annotated OBP and CSP of genomes of the triatome *R. prolixus* (Mesquita et al., International Consortium, 2015), the louse (*Pediculus humanus*; Kirkness et al., 2010) and the pea aphid (*Acyrthosiphon pisum*; Zhou et al., 2010). We added translated sequences from EST of two bugs from the Miridae family: *Adelphocoris lineolatus* (Gu et al., 2012, 2011; Sun et al., 2015) and *Apolygus lucorum* (Hua et al., 2013, 2012; Ji et al., 2013). Signal peptides of these selected protein sequences and of our translated contigs were deleted before alignment using MAFFT version 7 (Katoh et al., 2002) with the following options: E-INS-I with the BLOSUM62 scoring matrix and offset 0.1. We inspected the alignment by eye with BioEdit (Hall, 1999), and removed major gaps or sequences that were too short. We used ProtTest v3.2 (Darriba et al., 2011) to predict the best-fit models for protein evolution. The LG model with estimated Gamma parameter was retained. The tree was built with PhyML v3.0 (Guindon et al., 2010) with 100 bootstrap replicates and represented using iTOL web server (Letunic and Bork, 2007).

Mapping and counts

Much less RNA was extracted from the PMD sample than from others, resulting in a poor quality cDNA library. Consequently, this sample was excluded from the differential expression analysis and clustering study. Reads from other samples were mapped to the reference transcriptome after removing sequence adapters, using BWA with default options (Li and Durbin, 2009). Though we selected only one isoform per “gene” (the isoform states were assigned by Trinity), isoforms remained in our reference transcriptome. To avoid a resulting bias in our count analysis, we excluded reads with multiple hits. Count tables of reads were performed for each sample using SAMtools (Li et al., 2009). To evaluate the variation between technical replicates and sequencing runs, we proceeded to an ACP based on normalized counts with an rlog function of DESeq2. The ACP revealed high similarity between runs and technical replicates (Supplementary materials Figure 1) allowing them to be pooled in the subsequent analyses.

Differential expression analysis

Two different statistical methods were used for differential expression analysis: DESeq2 package v. 1.6.3 (Love et al., 2014) and EdgeR v. 3.8.5 (McCarthy et al., 2012). For both packages, we filtered contigs with a low coverage using HTSfilter (Rau et al., 2013). We used

the DESeq normalization method and a threshold of five mapped reads on average. DESeq2 automatically evaluates normalization and dispersion but we turned off the independent filtering. We chose the RLE model for normalization and a robust estimation method to evaluate dispersion with EdgeR. In both cases, count data was modeled with a negative binomial distribution in a Generalized Linear Models (GLM) with two factors: environment and sex. For the both methods, we controlled the factor “sex” when testing the factor “environment”. We compared the differential expression for the three modalities (sylvatic, peridomiciliary, domiciliary) of the “environment” factor and for the two modalities (male, female) of the “sex” factor. We applied the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) for multiple tests. The contigs were considered to be differentially expressed (DE) when the P_{adj} was below 0.05. We selected contigs that were differentially expressed with both the EdgeR and DESeq2 methods. These contigs were annotated using blastx (BLAST 2.2.29+) with the non-redundant protein database (version May 2015) and we selected only the best matches that had an e-value below 10^{-6} . The heatmap was built from chemosensory contig expression with DESeq2.

Gene clustering based on transcriptomic data

Clusters of co-expressed contigs were searched using HTScluster (Rau et al., 2015). This R package implements a Poisson mixture model to cluster expression observations. Despite the use of TMM normalization, HTScluster is sensitive to library size. The sample DFB was excluded from the analysis because of the lack of a technical replicate for this sample—otherwise there would have been one quarter fewer reads mapped than other samples for which technical replicates were collapsed. The count table was first filtered to remove contigs with a normalized average below 5. We selected the slope heuristic DDSE method. The distribution of contig expression in *T. brasiliensis* according to sex and environment for the clusters defined by HTScluster using a boxplot of the logarithm of normalized count is shown (figure 6).

Results

Reference transcriptome

We selected only the biggest isoforms from Trinity “genes” in the assembly 10c workflow established by Marchant et al., (2015a), and obtained a reference transcriptome of 48,290 contigs with an N50 of around 1,160 bp and a total length of 56,014,905 bp. The transcriptome appeared to be quite complete, with a CEGMA value of 94.35%. Males and females sampled from several populations were pooled to generate the 454 sequencing data so that the reference transcriptome represented all environmental conditions and sexes.

Annotation in reference transcriptome

We obtained 25 and 16 contigs annotated as CSP and OBP (Supplementary materials Table 1) for the *T. brasiliensis* transcriptome. All CSP contigs showed α helices, a signal peptide and four aligned cysteins (except for TbraCSP16). OBP contigs were less conserved. Nevertheless, the majority showed α helices, a signal peptide, and six aligned cysteins. To detect chemosensory genes in our samples, we looked at how many mapped reads in contigs were annotated as OBPs or CSPs. All chemosensory contigs presented mapped reads for all samples (except for TbraOBP12 in the SMC sample), suggesting they are expressed in both sexes and under all environmental conditions.

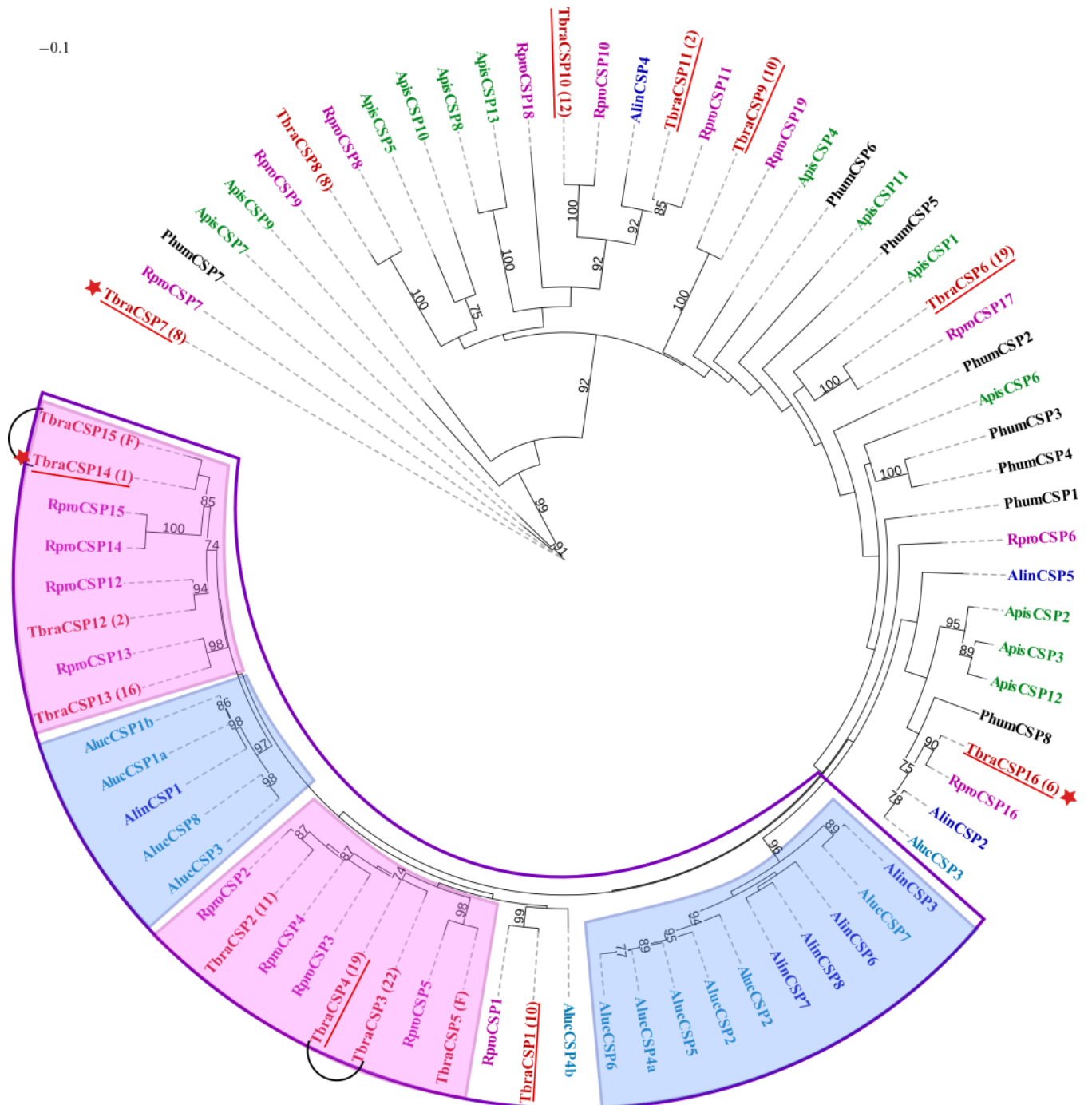
We retrieved available genomic data for other Paraneoptera species: 19 OBP and 13 CSP genes for *A. pisum* and 5 OBP and 8 CSP genes for *P. humanus* from the GenBank database; 27 OBP and 19 CSP genes for *R. prolixus* from Marchant et al. (2015b); and 16 OBP and 12 CSP sequences for *A. lucorum* and 13 OBP and 8 CSP sequences for *A. lineolatus* from the GenBank translation of antennae cDNA library.

Most of the OBPs and CSPs for *T. brasiliensis* had well supported orthologs in the *R. prolixus* genome, except for TbraOBP5 and TbraOBP7 (Figure 1). Some *R. prolixus* genes, though, had no tangible orthologs in the *T. brasiliensis* chemosensory transcriptome (=RproCSP6, RproCSP9, RproCSP18, RproOBP5, RproOBP7, RproOBP8, RproOBP15, RproOBP16, RproOBP19, RproOBP25). For five OBPs and four CSPs, several contigs of *T. brasiliensis* are grouped with a single *R. prolixus* gene (for example, TbraOBP23 and TbraOBP24 were clustered with RproOBP24, see clades joined with a brace in Figure 1). These contigs could be either paralogs that arose from gene duplication specific to *T. brasiliensis* (for example, TbraOBP15, 16 and 17) or orthologs that arose from an ancestral gene duplication in Triatominae (for example, TbraCSP3 and 4 were close to RproCSP3 and 4, two paralogs of *R. prolixus*). Nevertheless, as the *T. brasiliensis* chemosensory transcriptome was built from RNA pooled from several individuals, these contigs could also be allelic variants or isoforms from the same gene. Therefore, the chemosensory genes expressed in antennae and rostra amounts to 19 OBPs and 14 CSPs, if the putative variants or isoforms are removed or 25 OBPs and 16 CSPs, if all contigs were considered as unique genes. Hereafter, we will only reference the lower and more stringent estimate of gene number. Clear OBP and CSP expansions can be seen in the phylogeny of bloodsucking bugs (see pink blocks Figures 1 and 2) and more generally in Hemiptera (purple blocks, Figures 1 and 2) compared to the other species of Paraneoptera.

It is difficult to compare with the OBP and CSP gene repertoires of *A. lucorum* and *A. lineolatus* because data from both species originate from EST sequencing. However, both Triatomines clearly show more OBP and CSP genes than the Paraneoptera for which genomic data is available (*A. pisum*, *P. humanus*).

We also annotated contigs that code for the following: takeout proteins involved in feeding-relevant activities and in male courtship behavior (n=25), proteins involved in

protection against toxin penetration or in detoxification-like cuticle proteins (n=49), cytochrome P450s (n=72), and glutathione S-transferase (n=5) (Supplementary materials Table 2).



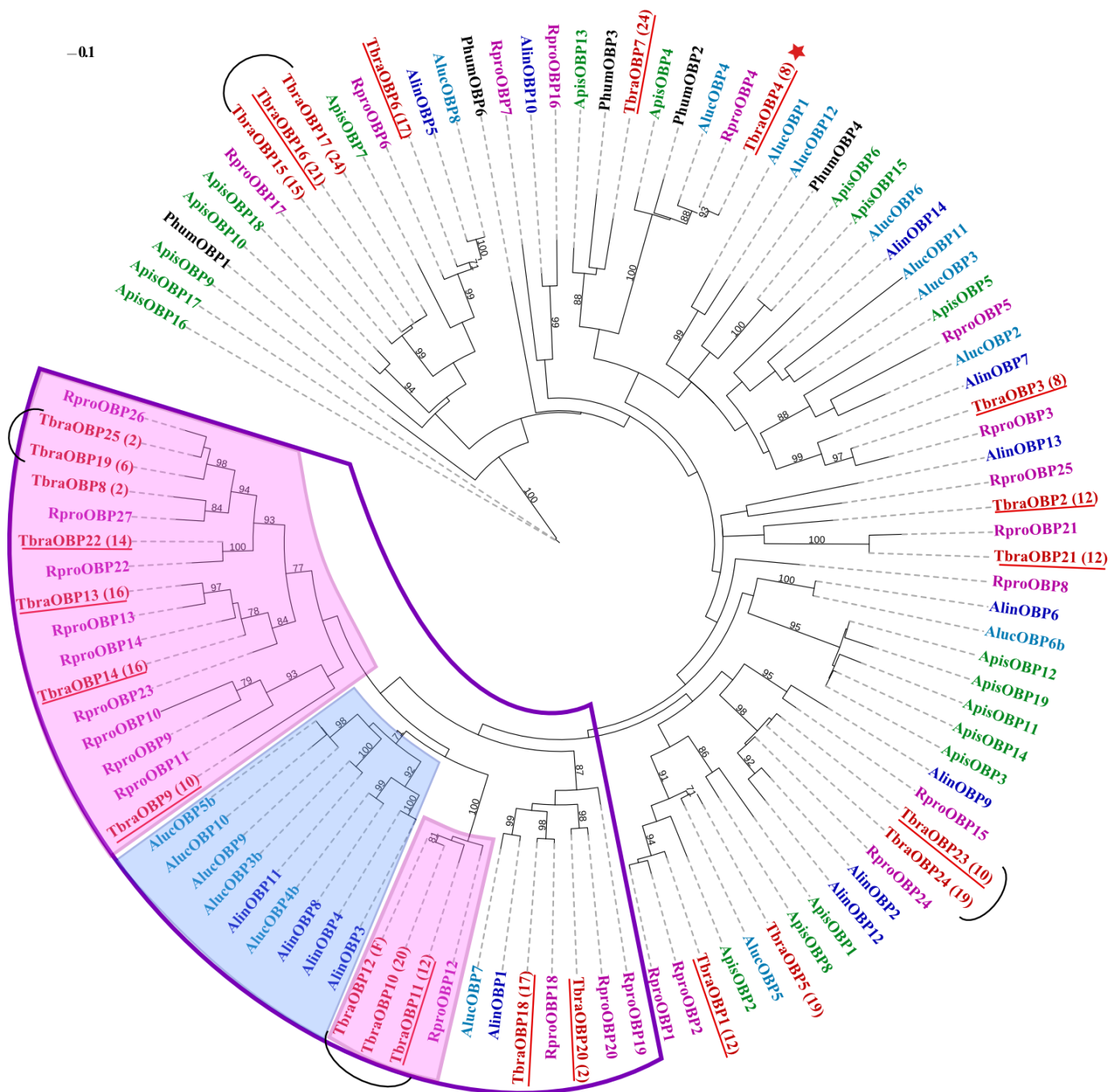


Figure 1 : Phylogenetic tree of CSP and OBP in Paraneoptera

Maximum likelihood trees of a) CSP and b) OBP in Paraneoptera: *Rhodnius prolixus* (pink) *Triatoma brasiliensis* (red), *Adelphocoris lineolatus* (dark blue), *Apolygus lucorum* (light blue), *Pediculus humanus* (black) and *Acyrtosiphon pisum* (green). Bootstrap >70% are indicated. *T. brasiliensis* contigs that are differentially expressed between environmental conditions are underlined. *T. brasiliensis* contigs that are differentially expressed between sexes are indicated with a star. Braces represent clades of several *T. brasiliensis* contigs that are grouped with a single predicted *R. prolixus* transcript, indicating different transcripts that arose either from polymorphism, alternative splicing of a single gene or paralogs. Boxes separate different clades and sub-clades (empty purple: Heteroptera, blue: Mirididae, pink: Triatominae). The numbers in parentheses indicate to which cluster contigs belong. (F) means that the probability that a contig belongs to a cluster is less than 0.9.

Sequencing, mapping and counts

The sequencing depth varied a lot between samples: from 22,692,898 (DFB) to 94,201,052 reads (SFC), excluding the cDNA library of the PMD sample (Table 1). The library size (total number of mapped reads), which depends both on sequencing depth and on the percentage of mapped reads, also varied between samples. For example, only 14.88% of DFB reads were mapped compared to 40.91% of SMC. Overall, the library size included between 3,376,263 and 27,848,905 mapped reads.

Most of the mapped reads were captured by very few genes, which is quite common in RNAseq analysis. The maximum percent of reads mapped to a single contig varied from 13.51% (PFD) to 46.26% (SMC) (Table 1). The most expressed contig was the same in all samples except DFB. It was associated by blast with a putative uncharacterized protein (gi:133916482) of the immune-related transcriptome of *Thermobia domestica* (Altincicek and Vilcinskis, 2007).

Differential expression analysis

We used both DESeq2 and EdgeR software, which differ in dispersion estimation methods. For the both software, dispersion decreases when expression increases, which is typical for RNAseq data among biological replicates. We used a two-factor general linear model (GLM) with both packages, allowing differential analysis despite the lack of biological replicates for some conditions. DESeq2 and EdgeR presented similar p-value distributions for all but the comparison between the sylvatic and peridomiciliary habitats. The peak was close to 0, corresponding to differentially expressed contigs between modalities and a uniform distribution for larger p-values, indicating a good fit to models generated by both types of software (Figure 3). However, the number of DE contigs differed between the two types of software. To increase the reliability of our results, we selected contigs in which we found DE with both EdgeR and DESeq2 (Table 2). We found numerous DE contigs (n=148) in both males and females (Table 2). Because peridomiciliary and domiciliary environments are anthroposystems, we expected more similarity between populations from these two ecotypes than between populations from sylvatic ecotopes. Surprisingly, a more DE contigs (n= 3875) were found between peridomiciliary and domiciliary populations than between sylvatic and domiciliary populations (n=29) and between sylvatic and peridomiciliary populations (n=29). Among DE contigs, most were over-expressed in the chemosensory organs of the peridomiciliary bugs, while they were mainly under-expressed in those of the domiciliary bugs.

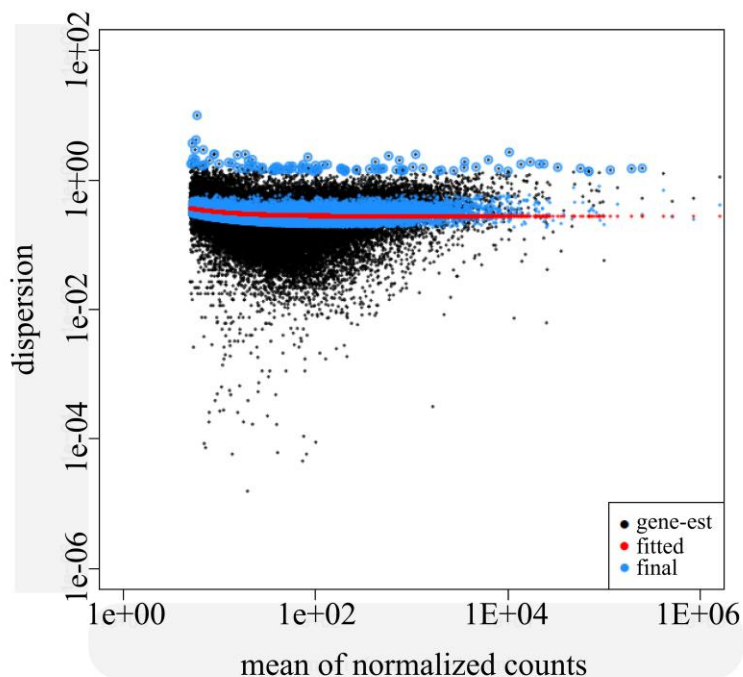


Figure 2: Dispersion estimates from DESeq2 and gene count measures between samples.

Dispersion of each gene (black), the trend line for all samples (red), the corrected value of the dispersion (blue) and outliers (black dot surrounded in blue). Variance decreases with the number of reads per contig until it stabilizes.

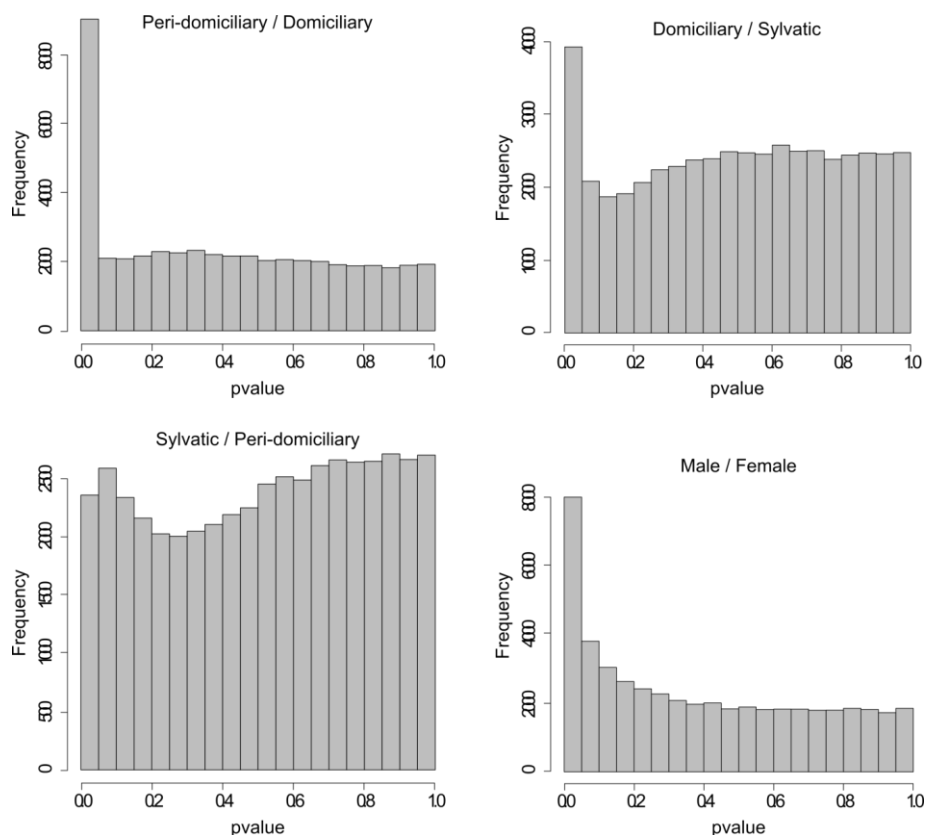


Figure 3: P-values distribution of differential expression analysis (DESeq2) per comparison

Conditions comparisons	DESeq2			EdgeR			Found in both DESeq2 and EdgeR			
	+	-	tot	+	-	tot	+	-	tot	Interested genes
M/F	148	4	152	1537	636	2173	144	4	148	2 CSP; 1 OBP; 3 TO; 3 P450; 14 CP
S/P	5	26	31	20	383	403	3	26	29	2 CP
D/S	11	292	303	2	30	32	2	27	29	3 OBP; 4 P450
D/P	19	4062	4081	427	5756	7183	16	3859	3875	13 CSP; 17 OBP; 37 P450; 14 TO; 16 CP; 3 GST

Table 2: Results of differential expression analysis.

Results are indicated for DESeq2 and EdgeR. We selected contigs that both packages identified as differentially expressed. “+” indicates number of over-expressed contigs in the first modality listed in each condition comparison (example, 148 contigs are found as over-expressed in males compared with females using DESeq2). “-” indicates the number of under-expressed contigs. “tot” is the total number of differentially expressed contigs per condition comparison. We annotated differentially expressed contigs found in both DESeq2 and EdgeR and we indicated the number of Odorant Binding Proteins (OBP), Chemosensory Proteins (CSP), *takeout* proteins (TO), Cuticular Proteins (CP), glutathione S-transferase (GST) and Cytochrome P450 (P450). D = domiciliary, P = peridomiciliary and S = sylvatic; M = male and F = female.

Annotation of differentially expressed genes

We were only able to annotate 50% of the DE contigs with the non-redundant proteins database. Three CSPs (CSP14, 15 and 16) and one OBP (OBP4) were differentially expressed between sexes, and all were over-expressed in males. A significant proportion of OBP and CSP contigs were differentially expressed between environments, particularly for the peridomiciliary-domiciliary comparison (Supplementary materials Table 2). Most were under-expressed in the domiciliary samples. We used a heatmap to compare chemosensory contig expression (Figure 4), revealing two distinct groups: sylvatic and peridomiciliary samples together in one and domiciliary samples in the other. Surprisingly, the sylvatic SFC sample was more similar to the domiciliary samples than to other sylvatic samples. This could explain why few DE contigs were detected in the sylvatic *versus* domiciliary comparison.

Among differentially expressed contigs, we also found the following contigs that code for *takeout* proteins (n=16), cytochrome P450s (n=38), glutathione S-transferases (n=3) and cuticle proteins (n=22) (Supplementary materials Table 2). These genes were over-expressed in males compared to females and in peridomiciliary and sylvatic samples compared to domiciliary samples. Some cuticle proteins were under-expressed in sylvatic samples compared to peridomiciliary samples.

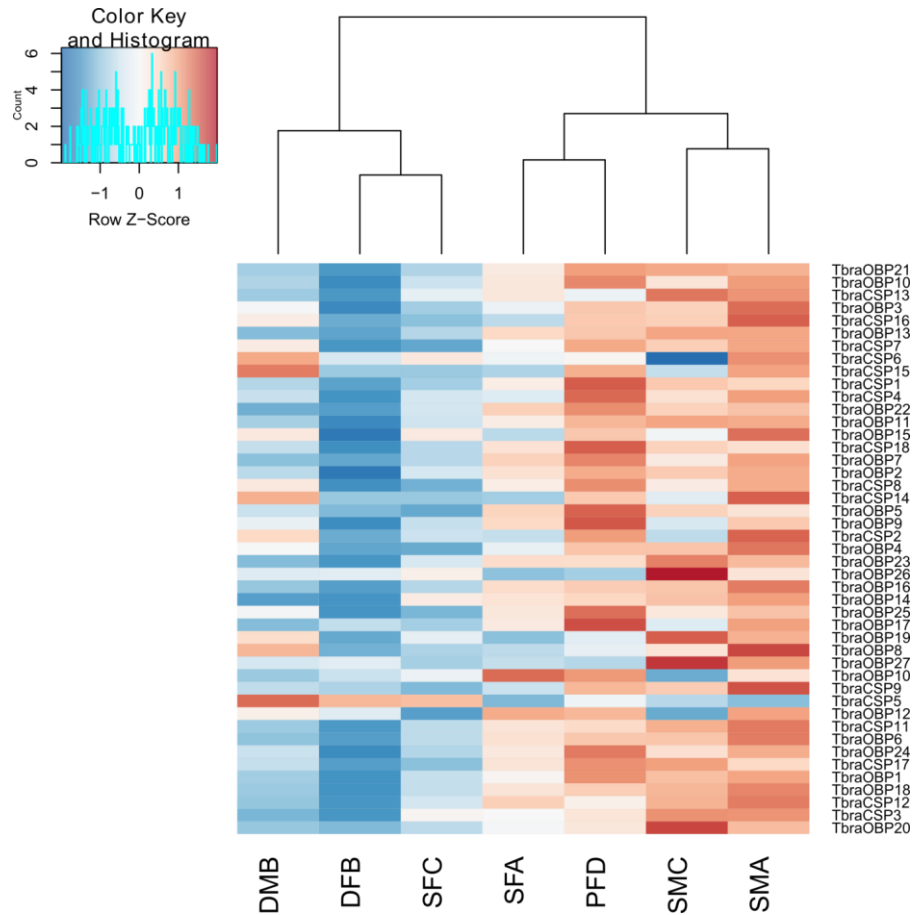


Figure 4: Heatmaps showing the sample distance of the expression data of differentially expressed chemosensory OBP/CSP contigs (DESeq2).

Gene Clustering

We selected the slope heuristic DDSE methods from HTScluster to estimate the number of clusters because BIC and ICL did not converge. We detected 24 clusters with between 28 and 14,385 contigs per cluster. Most contigs were reliably classified in a cluster with a maximum conditional probability close to 1 (Figure 5a). The quality of this classification, however, varied between clusters (Figure 5b). For example, only 21.65% of the contigs were classified in cluster 15 with a maximum conditional probability greater than 0.9, while 91.35% of contigs were classified in cluster 5 with the same probability. Clustering results agreed globally with results obtained from the differential expression study. For example, clusters 12 and 14 were made up primarily of contigs that were under-expressed in the domiciliary environment (Figure 6). Most contigs belonging to clusters 12 and 14 were significantly under-expressed in domiciliary/sylvatic or domiciliary/peridomiciliary comparisons with both DESeq2 and EdgeR.

Interestingly, clustering results may also provide more subtleties. For example, DE analysis revealed very few contigs that were over-expressed in females and in domiciliary samples. Nevertheless, clustering analysis revealed that contigs in clusters 3 and 20 showed higher expression in sylvatic females compared to all other samples (Figure 3). Similarly,

contigs in clusters 1, 5, 7, 13 and 23 showed higher expression in the domiciliary population than in the sylvatic/peridomiciliary populations (Figure 3). This apparent discrepancy may have arisen because over expression was too slight to be detected in the DE study. Joint analysis of several contigs with similar expression patterns would help to detect if this were the case. However, contigs with similar patterns of DE between modalities may be split into several clusters based on estimates of their expression pattern distribution. For example, this was the case for TbraOBP23 and TbraOBP24, which have very similar sequences (grouped with *Rhodnius prolixus* OBP24 in the phylogenetic tree) and are both over-expressed in the peridomiciliary population compared to the domiciliary population. They were, however, allocated to clusters 10 and 19, respectively. Similar conclusions can be drawn for TbraCSP3 and TbraCSP4.

Unfortunately, we were not able to test gene ontology enrichment of the cluster because of the low proportion of the transcriptome annotated (*T. brasiliensis* is quite distant from *R. prolixus* for which a reference genome is available).

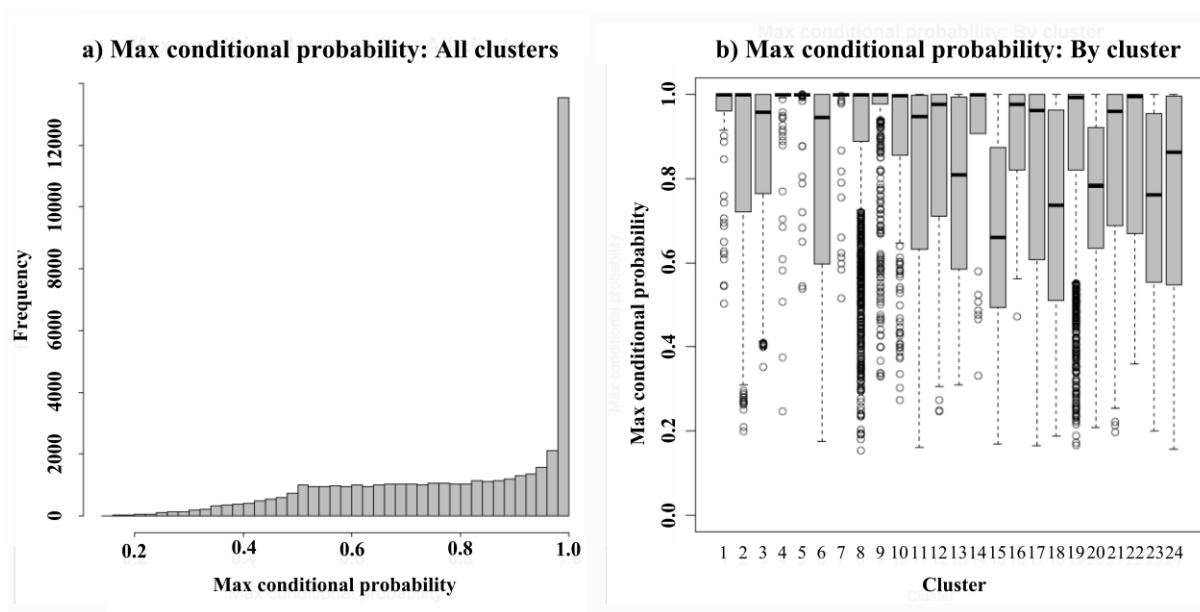


Figure 5: Histogram (a) of maximum conditional probabilities of cluster membership for all genes and (b) Boxplots of maximum conditional probabilities that genes assigned to each cluster (HTSCluster) are actually a member of that cluster.

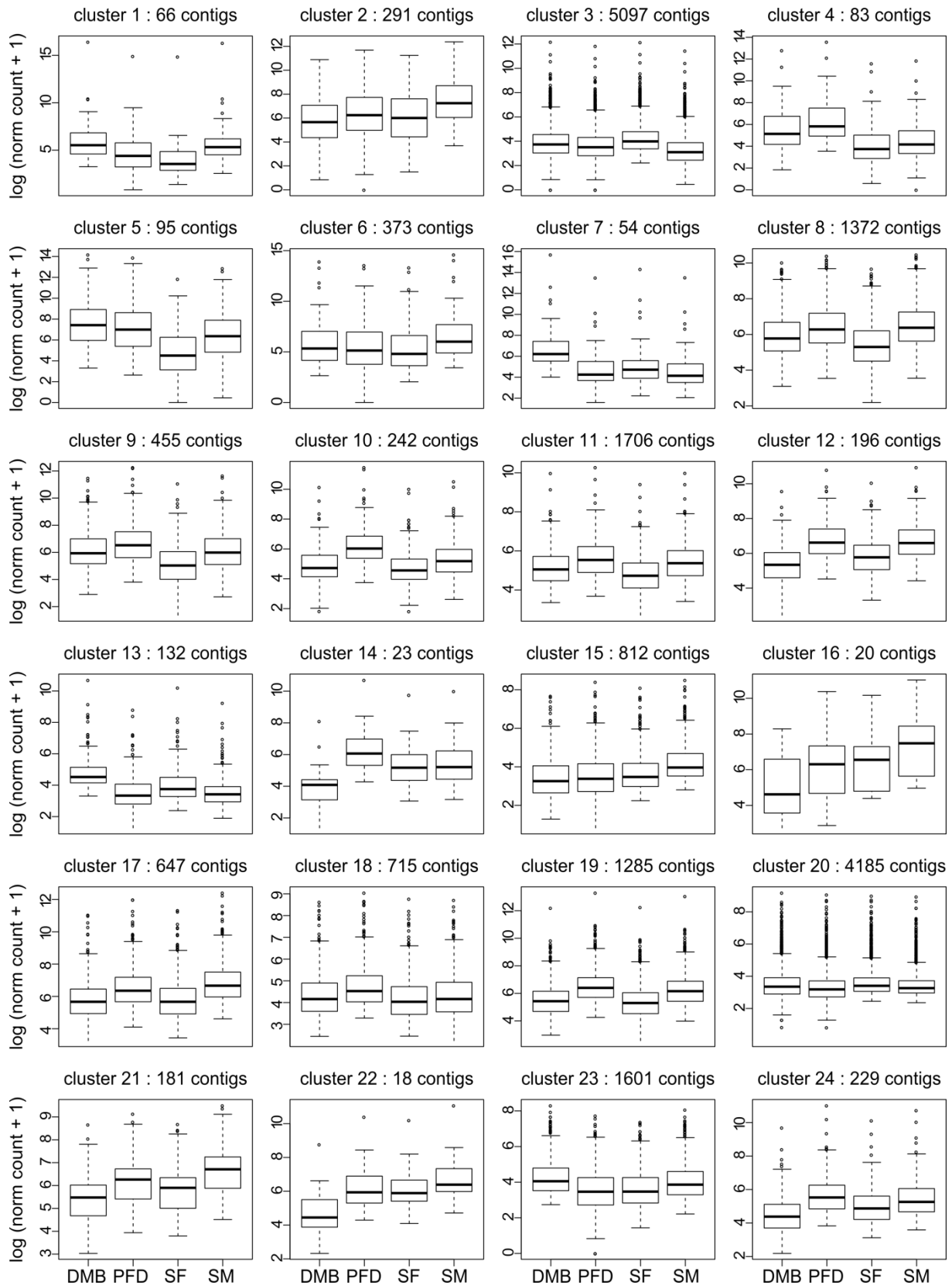


Figure 6: Distribution of contig expression in *T. brasiliensis* according to sex and environment for the 24 clusters defined by HTScluster.

Boxplot of the logarithm of normalized count data is shown. S = sylvatic, P = peridomiciliary and D = domiciliary. M = male and F = female.

Discussion

Reference transcriptome, CSP/OBP annotation

We estimated at least 19 OBP and 14 CSP genes in the *T. brasiliensis* genome. Compared to those annotated in the *R. prolixus* genome (27 OBP and 19 CSP: Marchant et al., 2015b), about 85% of the potential gene repertoire was retrieved from *T. brasiliensis*'s chemosensory transcriptome. Therefore, some CSPs and OBPs annotated in the *R. prolixus* genome seem to be lacking in the *T. brasiliensis* repertoire. Annotation of *T. brasiliensis* OBP/CSP genes might be incomplete because the transcriptome was performed from the head, antennae and rostrum while other OBPs and CSPs could be specifically expressed elsewhere, such as in the legs. Therefore the total number of CSP and OBP genes occurring in the *T. brasiliensis* genome is likely higher than we report here.

Loss/expansion events could also explain why there are fewer OBP/CSP genes detected in *T. brasiliensis* compared to *R. prolixus*, but we would need *T. brasiliensis* genomic data to evaluate this. Notably, two OBPs in *T. brasiliensis* have no orthologs in the *R. prolixus* genome. As they are clustered with other OBPs (*A. lucorum* and *P. humanus*), these copies could have been lost secondarily in *R. prolixus*. Bloodsucking bugs have more OBP and CSP genes than two other Paraneoptera: *A. pisum* (19 OBPs and 13 CSPs) and *P. humanus* (5 OBPs and 8 CSPs). How specialized a species is could be linked to how many OBP and CSP genes it has. The insect chemosensory genes are known to have evolved rapidly via gene duplication or loss events, in parallel with adaptation to new ecological niches (Guo and Kim, 2007; Kondrashov, 2012; Sánchez-Gracia et al., 2009; Vieira et al., 2007; Vieira and Rozas, 2011). The stable environment that results from a parasitic lifestyle could explain why *A. pisum* and *P. humanus* have fewer OBP and CSP genes (Kirkness et al., 2010; Zhou et al., 2010).

Two hypotheses might explain the OBP and CSP expansions in Chagas disease vectors that may be related to i) recruitment for a derived function, and more specifically in relation to hematophagy or ii) new chemosensory functions selected in response to environmental change. Striking similarities between OBPs and heme-binding proteins—like the occurrence of a signal peptide, six cysteines and a conserved PBP_GOBP domain (Mesquita et al., 2015)—support the hypothesis that OBP and CSP gene expansion is an adaptation to hematophagia. Some OBPs could have a derived function in providing protection from oxidative stress by binding the heme from digested blood hemoglobin, like heme-binding proteins (Dansa-Petretski et al., 1995). Moreover, some CSPs share an OS-D domain—a chemical characteristic of lipocalins—with genes expressed in the salivary transcriptome of another bloodsucking bug, *Triatoma matogrossensis*. Lipocalins are a class of proteins secreted in the saliva of bloodsucking bugs. They serve primarily as carriers of small ligands, and are used to disarm the host hemostatic machinery (Ribeiro et al., 2004). Some CSPs may play a similar role (Mesquita et al., in press).

Further studies are needed to explore such derived functions of olfactory/chemosensory proteins in bloodsucking bugs.

In Chagas disease vectors, gene expansion could also reflect a process of selecting for new chemosensory functions, providing an evolutionary advantage to individuals able to recognize various odor stimuli. Indeed, Triatominae are associated with diverse habitats, including nests or burrows of hosts, cavities of trees or palms, bromeliads or rocks or even anthropogenic habitats such as corrals, chickens coops or human dwellings. The Triatominae feed on a broad range of vertebrate hosts; 1150 host vertebrate species have been described (Patterson, 1994). Thus, OBP and CSP expansions could be explained as adaptation in blood-sucking insects that live in a putative diversity of ecological niches.

Differential gene expression

To maximize support for our results, we selected contigs that the EdgeR and DESeq2 techniques both identified as differentially expressed. Recent studies suggest that EdgeR and DESeq2 R packages are the best statistical methods for normalization and differential expression analysis of multi-factored experiments without a reference genome (Dillies et al., 2013, and Zhang et al., 2014). They are currently the only methods able to maintain a reasonable false-positive rate without decreasing the power of differentially expressed gene detection. They are stable under different sequencing depths. However, DESeq2 and EdgeR did not provide the same results with our dataset. EdgeR found more differentially expressed contigs than did DESeq2. Moreover, EdgeR found a higher log2-fold change than did DESeq2. Nevertheless, RNAseq is a recent technology and statistical research into how best to use differential expression analysis software with this kind of data is ongoing.

Recently, a gene clustering methodology based on transcriptomic data has been developed to highlight co-expressed contigs (HTSCluster, Rau et al., 2015). We found similar results using differential expression analysis and the HTSCluster gene clustering methodology. Furthermore, gene clustering provides additional detail about the dynamics of gene expression, especially in chemosensory contigs. We were able to annotate OBP or CSP contigs with very similar sequences to different clusters that reflect different expression patterns, suggesting their recent paralog or isoform status rather than the occurrence of allelic variants. This could support the idea that these sequences play a role in different functions—in other words, they could target different odorant molecules.

-Differential expression between sexes

148 contigs were differentially expressed in males and females. These were probably genes involved in sex-specific behavior or sexual dimorphism. Among these genes that were differentially expressed between sexes, we focused on those gene families that were also differentially expressed between environments. This isolated one OBP and three CSPs that were all over-expressed in males. Vitta et al. (2009) proposed that only male *T. brasiliensis* are attracted by sexual signals. OBPs and CSPs that are over-expressed in males in our study

could be involved in this kind of sexually dimorphic behavior. Previous studies of *A. lucorum* and *A. lineolatus*, two species that were included in the phylogenetic tree, showed that most of the OBPs and CSPs are over-expressed in female *A. lucorum* antennae. (Hua et al., 2012, 2013; Ji et al., 2013) while OBP or CSP over-expression may occur in either males or females in *A. lineolatus* (Gu et al., 2012, 2011; Sun et al., 2015). This suggests that expression and function of these proteins vary by Hemiptera species. But there could be a common function for AlucOBP4, which is closely related to TbraOBP4 in *T. brasiliensis* and is also over-expressed in males as in *A. lucorum* (Hua et al., 2012, 2013; Ji et al., 2013).

Three contigs annotated as *takeout* were also over-expressed in males. *Takeout* genes have been shown to affect male courtship (Sarav-Blat et al., 2000) and sexual dimorphism in locomotor activity, modulating the circulating juvenile hormone level in *Drosophila melanogaster* (Meunier et al., 2007). Locomotor activity also varies between sexes in *Triatoma infestans*, where females disperse more than males both by walking (Abraham et al., 2011) and flying (Minoli and Lazzari, 2006). In contrast, dispersal activity seems to be higher in male than in female *T. brasiliensis*: Sarquis et al., (2010) collected more males than females when using a flashlight to capture dispersing individuals. Differential courtship or sexual locomotor activity could explain the over-expression of takeout genes observed in *T. brasiliensis*.

We also detected differences between sexes in expression of contigs involved in insecticide resistance. The cytochrome P450 gene family is implicated in several functions other than detoxification. For example, in *Drosophila melanogaster*, a cytochrome P450 (cyp4d21) named *sxl* is differentially expressed between males and females and identified as a circadian-regulated gene involved in male courtship and mating success (Fujii et al., 2008). Differences in the ratio of cuticular proteins have also been found in the thoracic cuticle of male and female crickets (*Schistocerca gregaria* and *Locusta migratoria*). This difference has been linked to the specific mechanical ability of females to stretch intersegmental membranes (Andersen, 1998). Likewise, six genes that code for cuticular proteins were differentially expressed between the two sexes of *Anopheles gambiae*, which is probably linked to the sex-specific hematophagous behavior in this species (Togawa et al., 2008).

- Under-expression of contigs in domiciliary bugs

Our study revealed numerous differentially expressed genes in the transcriptomes of bugs sampled in different environments. This could either be explained by genetic variation between environmental populations after selection and adaptation to an anthropogenic system or by a phenotypic plasticity due to changes in gene expression induced by local environmental conditions. Whatever the mechanism, as the peridomiciliary environment is human-modified, we expected more similar phenotypes in the domiciliary and peridomiciliary populations (e.g. under- or over-expression of the same set of genes). Surprisingly, our comparison of peridomiciliary and domiciliary populations revealed much more differentially expressed contigs than did a comparison of sylvatic and domiciliary populations. Under-expression of the DE

genes for bugs from the domiciliary environment could reflect a less variable and more predictable biotope (abiotic and biotic conditions). A single sylvatic sample is probably responsible for the low number of DE genes detected between sylvatic and domiciliary populations in our study: patterns of expression in the sample SFC (females from sylvatic population C) were clearly more similar to those of domiciliary samples than to other sylvatic ones. The individuals that we sampled could have been domiciliary female migrants that dispersed either naturally or via human activity related to area surveillance by military guards. If this population came from domiciliary sources, the similar expression profile to the domiciliary population would support the genetic adaptation hypothesis, as a change in local environment did not induce a change in gene expression.

- Expression of chemosensory genes and takeout genes families

The DE contigs we detected were mainly under-expressed in the domiciliary samples. This general trend was also observed for contigs annotated as chemosensory genes OBP and CSP. Poor biotic diversity probably led to lower diversity of odorant stimuli in the domiciliary environment than in others. This reduction of stimuli could explain the under-expression of chemosensory genes in domiciliary bugs. Furthermore, several studies have linked chemosensory gene expression to feeding status. In the fly *Glossina morsitans*, transcription of CSP genes was related to host searching behavior (Liu et al., 2012). A recent study of bloodsucking bugs (Latorre Estivalis et al., 2015) revealed that, in *R. prolixus*, some odorant and ionotropic coreceptor genes are expressed less after a blood feeding. Similar results were previously found in *Anopheles gambiae* (Rinker et al., 2013). The nutritional status of sylvatic bugs varies throughout the year. Some authors correlate host searching with the dispersion process that may lead vectors to colonize domiciliary habitats (Lehane and Schofield, 1982; Sarquis et al., 2010; Schofield, 1980). In the domiciliary environment, putative human hosts are abundant in homes that—particularly in rural areas—can also include pets and other domestic animals. Therefore, domiciliary populations of *T. brasiliensis* have greater access to food than do other populations. Over-expression of chemosensory genes OBP and CSP in sylvatic and peridomiciliary bugs could be explained by the fact that they have higher foraging activity than do domiciliary bugs that under-express OBP and CSP genes.

Takeout proteins are also expressed in chemosensory organs and have been proposed to be associated with chemosensory perception in both taste and olfactory systems (Fujikawa et al., 2006). *Takeout* proteins link temporal cycle and feeding status information that are involved in the circadian cycle (Meunier et al., 2007; Sarov-Blat et al., 2000; So et al., 2000). Seven contigs were described as belonging to a *takeout* gene family in a transcriptome study of the digestive tract of *R. prolixus* (Ribeiro et al., 2014), but until now no *takeout* proteins had been annotated in *Triatoma*. We found a high number of *takeout* proteins in *T. brasiliensis* that were differentially expressed between the environments. *T. brasiliensis* has been shown to have a circadian cycle (Sarquis et al., 2010) much like that of *T. infestans* (Lazzari, 1992;

Barrozo et al., 2004). However, the circadian rhythm appears globally similar for sylvatic, peridomiciliary and domiciliary bugs because all are nocturnal. Therefore, under-expression of *takeout* genes in domiciliary bugs is probably linked to other functions attributed to this gene family. *Takeout* genes are also known to be involved in feeding-related metabolism and behavior (Sarov-Blat et al., 2000; So et al., 2000) and in the regulation of locomotor activity during foraging (Meunier et al., 2007). *Takeout* genes are induced by starvation, so it could be that they are over-expressed in sylvatic/peridomiciliary populations and not in domiciliary populations because of differences in feeding status.

-Expression of genes involved in insecticide resistance

Regular insecticide spraying campaigns have been used against triatomine bugs since the 1970s (Dias et al., 2002). Although, all *T. brasiliensis* tested until now have been highly susceptible to deltamethrin, an alpha-cyano-substituted pyrethroid insecticide (Pessoa et al., 2015), insecticide resistance has been demonstrated for some Chagas disease vectors, including *Triatoma infestans* (González Audino et al., 2004; Mougabure-Cueto and Picollo, 2015; Orihuela et al., 2008). We found DE genes involved in the insecticide resistance process, including genes coding the cytochrome P450, glutathione S-transferases and cuticle proteins. Indeed, several studies of cytochrome P450, the most well-studied gene family, has revealed over-expression in field populations of insects that are resistant to insecticide (Bergé et al., 1998; Boonsuepsakul et al., 2008; Cariño et al., 1994; Nkya et al., 2013; Scott, 1999; Xu et al., 2013). Cytochrome P450 genes are involved in detoxification. They catalyze various reactions; in particular, mono-oxygenation (Nkya et al., 2013; Scott, 1999). Nonetheless, the role of glutathione S-transferases in detoxification is still poorly understood. Several studies showed that they are linked to insecticide resistance and are over-expressed in pyrethroid-resistant insect populations (Enayati et al., 2005; Marcombe et al., 2012; Nkya et al., 2013; Vontas et al., 2001).

Insecticide resistance could also come from an increased ability to prevent toxins from penetrating the cuticle. Thus, changes to cuticle conformation or thickness could increase insecticide resistance (Gunning et al., 1995; Lin et al., 2012; Nkya et al., 2013). For example, over-transcription in a resistant strain of *An. gambiae* has suggested that the cuticle plays a role in response to selection pressures resulting from insecticides treatments (David et al., 2014). Cytochrome P450, glutathione S-transferases and cuticular proteins are all under-expressed in domiciliary bugs, which should be the most exposed to insecticide treatment. However, the domiciliary population was sampled in a house that had not been sprayed with insecticide for three years. Over-expression of these same genes in sylvatic and peridomiciliary bugs could reflect the involvement of detoxification enzymes in a large range of xenobiotics, including some that may differ according to the habitat (Nkya et al., 2013). In addition to their role in insecticide resistance, cuticle proteins also play a role in the response to environmental stresses (Zhang et al., 2008). The domiciliary environment probably has lower xenobiotic

diversity and more stable abiotic parameters than do sylvatic and peridomiciliary environments. This could explain why both detoxification and cuticle protein genes are expressed less in domiciliary bugs.

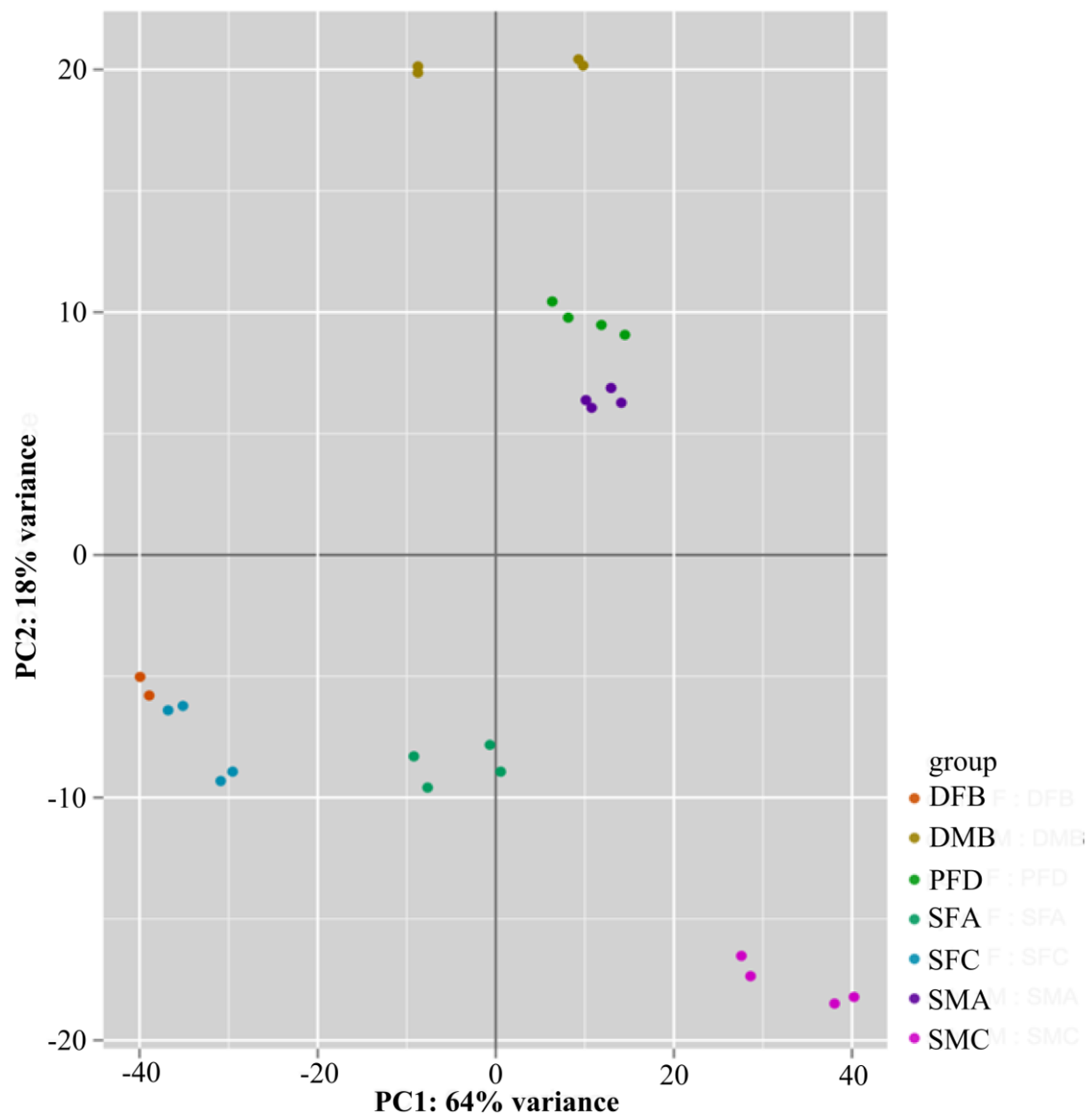
Conclusion

This study is the first analysis of differential expression based on RNAseq in *Triatoma brasiliensis*. Our aim was to better understand the domiciliation process. We focused on the chemosensory transcriptome because chemosensory genes are implicated in adaptation to changing environments. In the absence of a reference genome for *T. brasiliensis*, we used the methods of Marchant et al., (2015a) to build a reference transcriptome. RNAseq is a recent technology, so differential expression analysis software is still not optimized. Therefore, to generate results with strong support, we selected congruent results from two different programs: EdgeR and DESeq2. We then processed an independent clustering analysis based on the same expression pattern data. We characterized numerous DE transcripts from bugs of different environments and both sexes. The domiciliary population showed different expression profiles than those of sylvatic and peridomiciliary populations. Although, few samples from domiciliary and peridomiciliary populations were analyzed, making it difficult to generalize our conclusions, this study has identified new questions that need to be answered to understand the domiciliation process. Focusing on specific gene families, we highlighted the link between the environment from which a bug originates and expression of chemosensory genes (OBP, CSP), detoxification, cuticle proteins and *takeout* genes. These could be good candidate genes for playing a role in adaptation or in plastic rearrangement in response to environmental changes, which could include anthropic pressure. Further studies are needed to propose scenarios of molecular evolution of multigenic gene families (paralogs, isoforms, allelic variants) that will require new data—especially genomics. But understanding the molecular basis of vector adaptation to human dwellings creates the potential to develop new tools for disease vector control, perhaps by disrupting chemical communication.

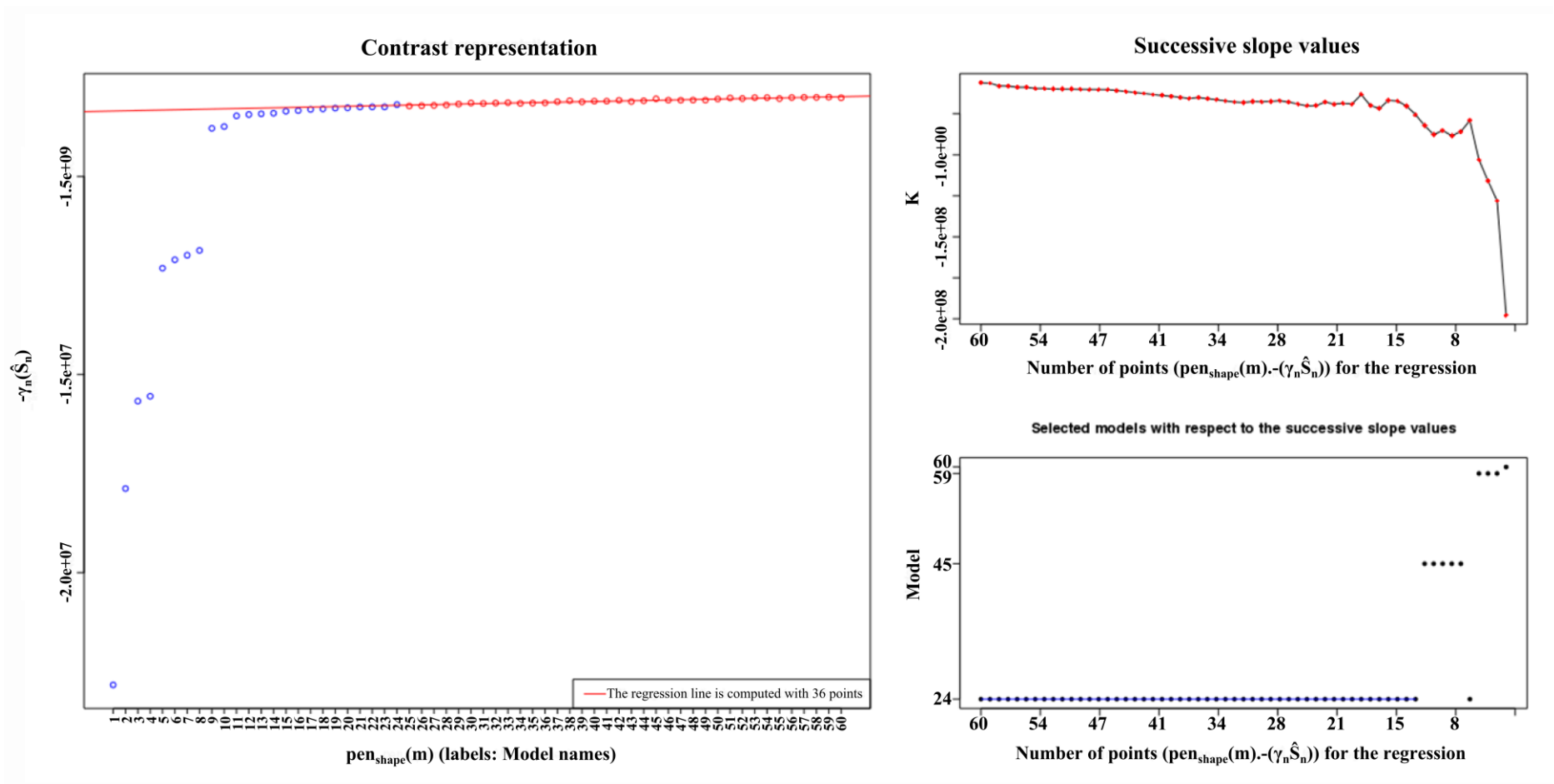
Acknowledgments

We would like to thank Andrea Rau (GABI, INRA, France) and Marie-Laure Martin-Magniette (IPS2, INRA Paris-Saclay), who help us to develop the statistical strategy of differential expression analysis and gene clustering. We thank Marie-Christine François (iEES, INRA Versailles, France) for help with the *T. brasiliensis* RNA extractions. The authors are also very grateful to the engineers of the bioinformatics platforms Genouest (Université de Rennes 1) and eBio (Université Paris Sud) for technical support. We also thank Sharilynn Wardrop for the comprehensive review of the English text. This study was funded by the ANR (ADAPTANTHROP project, ANR-097-PEXT-009) and supported by LabEx BASC (Université Paris Saclay, France). A. Marchant was funded by the Idex Paris Saclay, France and by ANR (ABC-PaPogen).

Supplementary



Supplementary figure 1: ACP based on normalized counts with the rlog function of DESeq2



Supplementary figure 2: Diagnostic plots provided by the *capushe* package for the DDSE approach of HTScluster

Supplementary materials Tables

Name	Contig	Blast NR ID	Annotation	%ident	Score	E-value	Cysteins	Psignal	α Helice	DE in comparison	Log2fold		Padj		HTScluster
											DESeq2	EdgeR	DESeq2	EdgeR	
TbraCSP1	Contig8299	gi 349592318	chemosensory protein 3 [Apolygus lucorum]	59.52	163	3.00E-45	4	19 and 20	8	D/P	-4.066	-5.629	0.001	0.004	10
TbraCSP2	Contig13506	gi 349592314	chemosensory protein 1 [Apolygus lucorum]	62.07	133	8.00E-35	4	18 and 19	8						11 (P<0.9)
TbraCSP3	Contig1318	gi 349592318	chemosensory protein 3 [Apolygus lucorum]	67.24	167	7.00E-45	4	20 and 21	8	D/P	-3.186	-3.761	0.003	0.023	22
TbraCSP4	Contig11307	gi 443908529	chemosensory protein 8 [Apolygus lucorum]	64.96	160	7.00E-47	4	18 and 19	8	D/P	-3.190	-3.929	0.004	0.007	19
TbraCSP5	Contig16746	gi 443908529	chemosensory protein 8 [Apolygus lucorum]	57.94	142	5.00E-38	4	18 and 19	8						filtered
TbraCSP6	Contig5540	gi 554894806	chemosensory protein 6 [Laodelphax striatella]	55.66	140	1.00E-35	4	18 and 19	8	D/P	-3.247	-4.828	0.013	0.019	19
TbraCSP7	Contig15930	gi 205326627	chemosensory protein [Apis cerana cerana]	65.05	142	2.00E-36	4	23 and 24	6	D/P	-2.571	-4.993	0.039	0.010	8
TbraCSP8	Contig13314	gi 215254086	putative chemosensory protein CSP9 [Nilaparvata lugens]	40.38	97.8	1.00E-20	4	22 and 23	8						8 (P<0.9)
TbraCSP9	Contig4898	gi 669092314	chemosensory proteins [Dendrolimus kikuchii]	38.79	85.5	9.00E-17	4	22 and 23	7	D/P	-3.060	-3.952	0.012	0.007	10
TbraCSP10	Contig5394	gi 270000350	chemosensory protein 4 [Adelphocoris lineolatus]	46.6	106	5.00E-25	4	22 and 23	8	D/P	-3.535	-4.801	0.004	0.007	12
TbraCSP11	Supercontig_454_29433	gi 270000350	chemosensory protein 4 [.]	56.69	156	4.00E-42	4	18 and 19	8	D/P	-3.521	-4.719	0.004	0.014	2
TbraCSP12	Contig3848	gi 349592318	chemosensory protein 3 [Apolygus lucorum]	59.68	165	7.00E-46	4	20 and 21	8	D/P	-2.945	-3.802	0.014	0.033	2
TbraCSP13	Contig18893	gi 349592314	chemosensory protein 1 [Apolygus lucorum]	66.35	153	7.00E-44	4	18 and 19	8	D/P	-2.556	-3.132	0.019	0.028	16
TbraCSP14	Contig9262	gi 349592318	chemosensory protein 3 [Apolygus lucorum]	57.26	152	8.00E-44	4	18 and 19	8	D/P M/F	-3.027 3.803	-5.409 5.707	0.020 1.849E-07	0.004 6.822E-06	1
TbraCSP15	comp176536	gi 349592318	chemosensory protein 3 [Apolygus lucorum]	60	131	1.00E-34	4	18 and 19	8	D/P M/F	-2.483 2.988	-3.707 4.636	0.032 3.916E-04	0.018 0.005	filtered
TbraCSP16	Contig2436	gi 443908519	chemosensory protein 3 [Apolygus lucorum]	80	193	5.00E-56	4	NO	6	D/P M/F	-3.544 3.019	-5.148 4.051	0.004 1.028E-05	0.005 9.745E-04	6
TbraOBP1	Contig6152	gi 33518701	odorant-binding protein precursor [Rhodnius]	80.14	244	3.00E-78	6	19 and 20	9	D/P	-3.692	-4.642	0.001	0.010	12

			prolixus]												
TbraOBP2	Contig6523	gi 23320749	odorant-binding protein 2 precursor [Zootermopsis nevadensis]	30	47	0.003	6	22 and 23	7	D/P	-2.966	-3.925	0.014	0.009	12 (P<0.9)
TbraOBP3	Contig6047	gi 573006064	odorant-binding protein 29 [Lygus lineolaris]	55.49	188	1.00E-55	6	23 and 24	9	D/P	-2.740	-4.003	0.022	0.019	8
TbraOBP4	Contig18371	gi 573006060	odorant-binding protein 27. partial [Lygus lineolaris]	70.59	254	3.00E-79	6	25 and 26	8	D/P M/F	-2.796 2.058	-4.034 2.828	0.022 0.045	0.006 0.009	8
TbraOBP5	comp179551_c0	gi 240247225	odorant-binding protein RproOBP6 precursor [Rhodnius prolixus]	90.34	280	1.00E-88	6	19 and 20	7						19 (P<0.9)
TbraOBP6	comp181497_c0	gi 573006002	odorant-binding protein 2 [Lygus lineolaris]	63.95	194	2.00E-58	6	26 and 27	8	D/P	-3.999	-5.373	0.001	0.008	17
TbraOBP7	Contig17590	gi 240247187	odorant-binding protein 13, partial [Acyrtosiphon pisum]	47.73	92	4.00E-19	6	NO	5	D/P	-2.811	-3.524	0.018	0.011	24
TbraOBP8	Contig14874	gi 700284171	putative odorant-binding protein 4 [Chinavia ubica]	52	43.1	0.028	4	19 and 20	8						2
TbraOBP9	Contig11714	gi 291195901	odorant-binding protein 1 [Apolygus lucorum]	26.09	58.5	2.00E-07	6	22 and 23	7	D/P	-2.495	-3.523	0.035	0.015	10
TbraOBP10	Supercontig_454_840	gi 586830420	odorant binding protein 8 [Adelphocoris suturalis]	30.28	67.8	4.00E-10	4	21 and 22	7						20 (P<0.9)
TbraOBP11	Contig11713	gi 573006042	odorant-binding protein 19 [Lygus lineolaris]	28.57	64.7	8.00E-10	4	21 and 22	8	D/P	-3.593	-4.674	0.002	0.004	12
TbraOBP12	Contig16927	gi 349592326	odorant binding protein 4 [Apolygus lucorum]	31	54.7	8.00E-06	4	NO	6						filtered
TbraOBP13	Contig11329	gi 240247221	odorant-binding protein RproOBP4 precursor [Rhodnius prolixus]	44.22	128	3.00E-33	6	18 and 19	7	D/P D/S	-4.390 -3.941	-5.885 -4.774	2.22E-04 2.38E-06	0.002 0.009	16
TbraOBP14	comp178106_c0	gi 240247221	odorant-binding protein RproOBP4 precursor [Rhodnius prolixus]	66.44	185	6.00E-55	6	17 and 18	8	D/P D/S	-3.705 -3.596	-4.976 -4.359	0.003 4.40E-05	0.001 0.002	16
TbraOBP15	Contig16558	gi 270000366	odorant-binding protein 4 [Adelphocoris lineolatus]	51.91	142	5.00E-37	6	22 and 23	7						15
TbraOBP16	Supercontig_454_5612	gi 270000366	odorant-binding protein 4 [Adelphocoris lineolatus]	51.91	142	5.00E-38	6	22 and 23	3	D/P	-2.329	-3.027	0.039	0.012	21
TbraOBP17	Contig16870	gi 700284175	putative odorant-binding protein 1 [Telenomus podisi]	56.34	87.8	5.00E-18	3	NO	5						24 (P<0.9)
TbraOBP18	Contig3641	gi 270000360	odorant-binding protein 1 [Adelphocoris lineolatus]	45.45	135	5.00E-36	6	20 and 21	6	D/P	-3.492	-4.584	0.003	0.013	17
TbraOBP19	Contig14074	gi 240247219	odorant-binding protein RproOBP2 precursor, partial [Rhodnius prolixus]	29.06	55.1	9E-07	4	NO	7						6

TbraOBP20	Contig19415	gi 270000360	odorant-binding protein 1 [Adelphocoris lineolatus]	40.56	130	3E-33	6	19 and 20	7	D/P	-3.643	-4.531	0.002	0.011	2
TbraOBP21	Contig13477	gi 573459726	putative odorant binding protein 77 [Nasonia vitripennis]	32	43.1	0.033	6	30 and 31	6	D/P	-3.605	-4.593	0.002	0.006	12
TbraOBP22	Contig9624	gi 240247219	odorant-binding protein RproOBP2 precursor. partial [Rhodnius prolixus]	34.19	85.1	1.00E-16	6	19 and 20	7	D/P D/S	-4.239 -3.408	-5.278 -3.944	1.85E-4 1.57E-05	0.001 0.019	14
TbraOBP23	Contig14361	gi 270000474	odorant binding protein 12 [Adelphocoris lineolatus]	59.57	66.6	3.00E-10	3	21 and 22	3	D/P	-2.779	-4.325	0.024	0.006	10 (P<0.9)
TbraOBP24	Contig16524	gi 573006006	odorant-binding protein 4 [Lygus lineolaris]	55.93	160	1.00E-43	6	21 and 22	6	D/P	-3.191	-4.640	0.012	0.023	19
TbraOBP25	Contig12789	gi 240247219	odorant-binding protein RproOBP2 precursor, partial [Rhodnius prolixus]	30.77	68.9	3.00E-11	4	19 and 20	9						2

Supplementary Table 1: Blast annotation of CSP and OBP, and information about differential expression status.

Results of Blast: For all contigs that matched with a CSP or an OBP from the non-redundant protein database, we indicate the ID of this protein (Blast NR ID) and its denomination (Annotation). For each match, we provide the percentage of identity (%ident), the score and the e-value. Each contig that matches with an OBP or CSP were translated and their protein analyzed by checking for the number of conserved cysteins, the presence of peptide signal and the number of α -helices. For each contig annotated as OBP or CSP, we indicated in which comparison both DESeq2 and EdgeR found it to be differentially expressed. D = domiciliary, P = peridomiciliary and S = sylvatic; M = male and F = female. The log2fold and the adjusted p-value for both packages are provided, as well as the cluster number allocated by HTScluster.

Contig	NR	Annotation	%id	Score	E-value	DE in comparison	Log2fold		Padj		HTS Cluster
							DESeq2	EdgeR	DESeq2	EdgeR	
comp171165_c2	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	36,5	181	2,00E-49	M/F	2.874	3.948	6.01E-04	5.18E-04	5
Contig9476	gi 389610927	takeout/JHBP like protein [Papilio polytes]	50,4	242	1,00E-73	D/P M/F	-2.607 2.290	-4.010 3.324	0.027 0.011	0.010 0.006	9
Contig3858	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	69,5	174	3,00E-70						9
Contig16387	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	43,1	209	2,00E-61	D/P M/F	-3.334 4.058	-4.845 5.182	0.008 2.77E-11	0.006 1.68E-05	2
Contig11061	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	37,3	181	1,00E-49	D/P	-2.929	-5.754	0.025	0.015	19
Contig242	gi 646720336	Protein takeout [Zootermopsis nevadensis]	39,8	207	1,00E-58	D/P	-2.960	-4.548	0.020	0.016	12
Contig4693	gi 646707272	Protein takeout [Zootermopsis nevadensis]	56,1	294	2,00E-92	D/P	-2.678	-4.487	0.023	0.031	8
Contig8939	gi 193645781	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	42	192	2,00E-54	D/P	-4.047	-5.422	9.34E-04	0.007	12
Contig15222	gi 326381129	takeout-like [Acyrtosiphon pisum]	40,8	211	2,00E-62	D/P	-3.940	-5.669	0.002	0.004	10
Contig12334	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	38,4	155	2,00E-41	D/P	-2.954	-4.061	0.016	0.010	12
comp158409_c0	gi 242022243	protein takeout precursor, putative [Pediculus humanus corporis]	30,1	79,7	2,00E-14	D/P	-2.758	-3.686	0.019	0.044	2
Contig1601	gi 242024397	protein takeout precursor, putative [Pediculus humanus corporis]	37.02	160	4e-41	D/P	-2.657	-4.261	0.029	0.019	19
Contig6740	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	43.33	214	4e-61	D/P	-2.981	-4.965	0.020	0.025	17
comp171070_c0	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	48.46	237	3e-66						21
Contig11584	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	41.38	211	2e-62						7
Supercontig_454_5739	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	48.46	237	5e-69						15
Contig17773	gi 242009643	protein takeout precursor, putative [Pediculus humanus corporis]	63.01	102	7e-24						21 (P<0.9)

comp158409_c0	gi 242022243	protein takeout precursor, putative [Pediculus humanus corporis]	30.08	79.7	2e-14	D/P	-2.758	-3.686	0.019	0.044	2
Contig2664	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	44.22	239	2e-70						9
mira_454_rep_c72681	gi 18140056	takeout [Aedes aegypti]	43.55	117	2e-27	D/P	2.980	8.346	0.030	0.004	17 (P<0.9)
comp126832_c0	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	51.54	246	3e-77						23
Contig645	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	38.31	187	5e-52	D/P	2.757	3.346	0.015	0.045	16
Contig6143	gi 326381129	takeout-like [Acyrtosiphon pisum]	39.58	112	1e-38						12
Supercontig_454_2845	gi 339765128	cytochrome P450 CYP417A1 [Nilaparvata lugens]	30,5	113	4,00E-25						6
comp174812_c0	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	29,9	179	7,00E-47						8
comp145626_c1	gi 501295299	cytochromeP450 [Riptortus pedestris]	36,1	311	2,00E-96						6
Supercontig_454_10385	gi 568259601	cytochrome P450 [Anopheles darlingi]	36,9	73,6	4,00E-15						12 (P<0.9)
comp171414_c0	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	29,3	203	4,00E-55	M/F	2.808	3.412	4.421E-05	9.00E-05	2
comp154816_c0	gi 339896273	cytochrome P450 [Bemisia tabaci]	50	191	1,00E-53						23 (P<0.9)
Contig11448	gi 501295299	cytochromeP450 [Riptortus pedestris]	44,1	194	2,00E-52	D/P	-2.757	-3.346	0.015	0.045	22 (P<0.9)
Supercontig_454_43924	gi 391346350	PREDICTED: cytochrome P450 3A31-like [Metaseiulus occidentalis]	32,7	92,4	1,00E-16						3
comp146998_c0	gi 31223075	cytochrome P450 CYP6Y1 [Anopheles gambiae]	39,3	225	1,00E-65						12 (P<0.9)
Supercontig_454_1667	gi 564227928	PREDICTED: cytochrome P450 4V2-like, partial [Alligator mississippiensis]	52,1	211	3,00E-63	D/P	-2.561	-3.304	0.020	0.011	17
Contig19271	gi 501295299	cytochromeP450 [Riptortus pedestris]	43,9	135	4,00E-45	D/P	-2.756	-3.586	0.021	0.005	10 (P<0.9)
Contig8480	gi 501295299	cytochromeP450 [Riptortus pedestris]	44,4	478	7,00E-159	D/P	-2.666	-3.472	0.022	0.028	17
Contig4340	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	40,2	179	6,00E-47	D/P	-2.953	-4.233	0.017	0.011	19
Contig18531	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	38,7	340	1,00E-102	D/P	-2.568	-3.306	0.022	0.024	17
Contig1415	gi 501295299	cytochromeP450 [Riptortus pedestris]	38,4	322	3,00E-98	D/P	-2.926	-4.367	0.019	0.015	17

newbler_454_04299	gi 510813177	cytochrome P450, partial [Laodelphax striatella]	48,4	139	8,00E-35	D/P	-2.532	-3.613	0.030	0.015	17
comp174285_c0	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	36	290	4,00E-87	D/P	-3.130	-4.069	0.010	0.004	19
Supercontig_454_7993	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	38,8	205	2,00E-57	D/P D/S	-3.743 -3.635	-4.849 -4.266	0.002 6.87E-06	0.007 0.045	2
Contig2856	gi 646716643	putative cytochrome P450 6a13 [Zootermopsis nevadensis]	31,4	177	4,00E-80	D/P	-3.545	-5.407	0.009	0.006	12
Contig9883	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	39,2	212	4,00E-59	D/P D/S	-3.743 -3.478	-4.849 -4.069	0.002 3.06E-05	0.007 0.043	2
comp175484_c9	gi 501295299	cytochromeP450 [Riptortus pedestris]	47,4	154	1,00E-39	D/P D/S	-3.774 -4.052	-5.112 -4.896	0.002 2,60E-07	0.005 0.009	16
Contig18029	gi 270006370	cytochrome P450 6BQ4 [Tribolium castaneum]	32,4	161	2,00E-64	D/P	-3.856	-5.108	0.002	0.007	17
Contig13659	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	35	211	1,00E-58	D/P	-3.648	-6.102	0.008	0.004	12
mira_454_rep_c18806	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	47,5	179	3,00E-48	D/P	-2.638	-3.276	0.021	0.006	21 (P<0.9)
Contig2511	gi 501295299	cytochromeP450 [Riptortus pedestris]	35,7	359	5,00E-112	D/P	-3.087	-5.167	0.020	0.013	2
comp172878_c4	gi 645004806	PREDICTED: cytochrome P450 4g15 [Nasonia vitripennis]	31,4	95,1	3,00E-18	D/P D/S	-3.172 -3.249	-4.229 -3.896	0.014 6.62E-04	0.008 0.043	16 (P<0.9)
comp175484_c3	gi 501295299	cytochromeP450 [Riptortus pedestris]	38	193	3,00E-49	D/P	-3.021	-4.084	0.017	0.019	21
comp120212_c0	gi 646707025	putative cytochrome P450 CYP44 [Zootermopsis nevadensis]	40.21	398	2,00E-129	M/F	1.918	2.235	0.023	0.010	1
comp126058_c1	gi 61611906	cytochrome P450 [Culex pipiens pallens]	34.28	315	8,00E-97	D/P	-2.597	-3.625	0.028	0.008	24
comp147212_c0	gi 662213714	PREDICTED: cytochrome P450 302a1, mitochondrial [Diaphorina citri]	45.79	199	4,00E-56						6 (P<0.9)
comp147602_c0	gi 60686423	cytochrome P450 [Mayetiola destructor]	33.85	126	5,00E-30						15 (P<0.9)
comp150255_c0	gi 501295299	cytochromeP450 [Riptortus pedestris]	34.62	198	5,00E-55	D/P	-2.717	-3.675	0.025	0.010	12
comp163607_c0	gi 291464089	cytochrome P450 4G4 [Manduca sexta]	65.91	732	0.0	D/P	-2.605	-3.252	0.020	0.014	17
comp163933_c1	gi 498928672	PREDICTED: cytochrome P450 4g1-like [Ceratitis capitata]	48.37	149	4,00E-37	D/P	-2.231	-2.928	0.048	0.019	21 (P<0.9)
comp165506_c0	gi 408724249	cytochrome P450 CYP18A1, partial [Laodelphax striatella]	69.88	364	6,00E-121						15 (P<0.9)
comp168063_c0	gi 380015389	PREDICTED: cytochrome P450 306a1-like, partial [Apis florea]	38.95	148	2,00E-38						18 (P<0.9)
comp168551_c0	gi 501295299	cytochromeP450 [Riptortus pedestris]	40.05	345	7,00E-110	D/P	-2.833	-3.537	0.015	0.008	21 (P<0.9)

comp170482_c1	gi 501295299	cytochromeP450 [Riptortus pedestris]	42.89	364	4,00E-117	D/P	- 2.314	-2.858	0.030	0.018	17 (P<0.9)
comp173258_c1	gi 383857845	PREDICTED: probable cytochrome P450 6a14-like [Megachile rotundata]	42.01	188	3,00E-52						11 (P<0.9)
comp173258_c2	gi 478734979	cytochrome P450 CYP6BQ22 [Dastarcus helophoroides]	47.01	238	1,00E-69						19
comp173921_c0	gi 408724329	cytochrome P450 CYP6AY3v2 [Laodelphax striatella]	32.72	279	3,00E-80						17 (P<0.9)
comp174285_c0	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	35.98	289	4,00E-87	D/P	-3.130	-4.069	0.011	0.004	19
comp174572_c3	gi 646720986	Cytochrome P450 4C1 [Zootermopsis nevadensis]	32.97	93.2	1,00E-16						21
comp175018_c6	gi 350397446	PREDICTED: probable cytochrome P450 301a1, mitochondrial-like isoform 1 [Bombus impatiens]	68.04	501	1,00E-170						8
comp175049_c2	gi 339896295	cytochrome P450 [Bemisia tabaci]	50.00	63.5	3,00E-08						24 (P<0.9)
comp175685_c2	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	26.85	147	2,00E-34						2
comp176576_c1	gi 408724329	cytochrome P450 CYP6AY3v2 [Laodelphax striatella]	51.00	513	1,00E-174						15 (P<0.9)
comp294098_c0	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	40.24	330	1,00E-103	D/P	-2.261	-2.866	0.034	0.011	10
comp55327_c0	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	39.31	341	4,00E-107	D/P	-2.584	-3.387	0.026	0.015	12
Contig10143	gi 501291494	cytochromeP450 [Riptortus pedestris]	54.48	166	4,00E-48	D/P	-2.475	-3.126	0.025	0.004	21 (P<0.9)
Contig10144	gi 646477539	putative cytochrome P450 6a13 [Zootermopsis nevadensis]	50.00	91.7	1,00E-20						filtered
Contig10638	gi 242006280	NADPH--cytochrome P450, putative [Pediculus humanus corporis]	93.94	68.9	1,00E-10						15 (P<0.9)
Contig12459	gi 19879426	cytochrome P450 monooxygenase CYP6X1v2 [Lygus lineolaris]	48.17	355	2,00E-110	D/P	-2.597	-3.950	0.030	0.028	8
Contig17074	gi 642938945	PREDICTED: probable cytochrome P450 6a23, partial [Tribolium castaneum]	45.77	142	1,00E-37						24 (P<0.9)
Contig18608	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	57.83	108	2,00E-23	D/P	-2.651	-3.485	0.020	0.019	21
Contig18738	gi 408724221	cytochrome P450 CYP417A2v2 [Laodelphax striatella]	34.73	229	5,00E-65	D/P	-2.534	-3.2264	0.023	0.007	12 (P<0.9)
Contig2982	gi 380293835	NADPH-cytochrome P450 reductase [Cimex lectularius]	83.46	1171	0.0						2
Contig3204	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	37.89	381	2,00E-119						2
Contig3760	gi 501295299	cytochromeP450 [Riptortus pedestris]	44.53	442	5,00E-146	D/P	2.891	3.646	0.013	0.034	21
Contig3961	gi 383860255	PREDICTED: cytochrome P450 6a2-like [Megachile rotundata]	35.02	140	4,00E-34						18 (P<0.9)
Contig4442	gi 646701016	Cytochrome P450 9e2 [Zootermopsis nevadensis]	38.69	372	4,00E-114						19

Contig6425	gi 408724255	cytochrome P450 CYP6BD10v2 [Laodelphax striatella]	42.34	361	1,00E-108	D/P	-2.424	-4.154	0.048	0.019	24
Contig8480	gi 501295299	cytochromeP450 [Riptortus pedestris]	44.40	478	7,00E-159	D/P	-2.666	-3.472	0.022	0.028	17
Contig9376	gi 501295299	cytochromeP450 [Riptortus pedestris]	40.17	260	7,00E-78						18 (P<0.9)
mira_454_rep_c4693	gi 662186779	PREDICTED: cytochrome P450 4g1-like [Diaphorina citri]	51.56	70.9	5,00E-12						Filtered
newbler_454_02565	gi 91084715	cytochrome P450 6BT1 [Tribolium castaneum]	50.00	91.3	1,00E-17						15 (P<0.9)
Supercontig_454_24630	gi 391346350	PREDICTED: cytochrome P450 3A31-like [Metaseiulus occidentalis]	36.96	45.8	1,00E-06						20 (P<0.9)
Supercontig_454_43655	gi 646720986	Cytochrome P450 4C1 [Zootermopsis nevadensis]	36.90	58.9	4,00E-06						22
Supercontig_454_46254	gi 339896301	cytochrome P450 [Bemisia tabaci]	39.34	149	5,00E-38	D/P	-2.958	-4.332	0.021	0.001	19 (P<0.9)
Supercontig_454_7517	gi 501295299	cytochromeP450 [Riptortus pedestris]	30.77	125	4,00E-29						16 (P<0.9)
Supercontig_454_829	gi 662186779	PREDICTED: cytochrome P450 4g1-like [Diaphorina citri]	51.56	70.9	5,00E-12						Filtered
comp161828_c0	gi 387413531	glutathione s-transferase T1 [Sogatella furcifera]	54,1	248	3,00E-77	D/P	-2.259	-3.096	0.041	0.022	8
Contig13312	gi 700894839	glutathione S-transferase [Diaphorina citri]	41,5	143	4,00E-38	D/P	-2.364	-3.574	0.042	0.017	8
Contig14872	gi 387413743	glutathione s-transferase Z1 [Sogatella furcifera]	89,4	414	1,00E-142	D/P	-2.892	-4.450	0.020	0.014	19
comp171728_c1	gi 641654330	PREDICTED: glutathione S-transferase theta-1-like [Acyrtosiphon pisum]	53,7	258	2,00E-81						21
comp175562_c1	gi 387413424	glutathione s-transferase M2 [Nilaparvata lugens]	66.67	204	6e-60						18
Supercontig_454_1650	gi 817068565	PREDICTED: larval cuticle protein A2B-like [Athalía rosae]	76,9	112	2,00E-26	M/F	3.402	4.184	1.79E-07	3.19E-04	2
Contig15155	gi 646706208	Larval cuticle protein A3A [Zootermopsis nevadensis]	88,3	116	8,00E-28	D/P M/F	-3.131 3.330	-4.999 4.704	0.017 7.78E-06	0.005 1.24E-04	2
Supercontig_454_44553	gi 817068565	PREDICTED: larval cuticle protein A2B-like [Athalía rosae]	76,9	110	2,00E-25	M/F	2.402	2.921	0.003	0.002	6
comp173141_c0	gi 117622	Cuticle protein 8 [Locusta migratoria]	81	107	9,00E-25	M/F	3.260	3.942	4.41E-06	1.24E-04	2
Contig5048	gi 817214821	PREDICTED: cuticle protein 21-like [Orussus abietinus]	75	132	1,00E-34	S/P D/P M/F	-3.040 -2.208 5.100	-4.400 -3.652 6.435	0.035 0.045 9.62E-19	0.003 0.017 5.05E-09	1
Supercontig_454_438	gi 817054670	PREDICTED: cuticle protein 19-like [Athalía rosae]	84,1	123	2,00E-29	D/P M/F	-2.864 2.964	-5.149 4.544	0.023 2.86E-04	0.019 0.008	8

Contig9755	gi 117622	Cuticle protein 8 [<i>Locusta migratoria</i>]	59,7	71,6	4,00E-11	D/P M/F	-3.141 1.905	-4.013 2.384	0.007 0.028	0.009 0.034	17
comp167139_c0	gi 501291136	cuticle protein, putative [<i>Riptortus pedestris</i>]	91,9	120	7,00E-29	D/P M/F	-3.359 3.319	-5.050 4.558	0.011 3.41E-06	0.004 1.26E-04	2
Contig7919	gi 646704183	Cuticle protein [<i>Zootermopsis nevadensis</i>]	62,5	152	1,00E-38	M/F	2.360	4.180	0.031	0.016	5
comp155262_c0	gi 646715957	Cuticle protein 7 [<i>Zootermopsis nevadensis</i>]	68,8	122	1,00E-28						filtered
Contig1532	gi 240848893	cuticular protein CPG12-like precursor [<i>Acyrtosiphon pisum</i>]	82,3	87,8	7,00E-16	D/P M/F	-3.304 2.163	-4.972 3.170	0.012 0.032	0.009 0.025	9
Supercontig_454_5892	gi 646715957	Cuticle protein 7 [<i>Zootermopsis nevadensis</i>]	68,8	122	2,00E-28	D/P S/P M/F	-3.395 -3.688 3.407	-6.019 -5.693 5.278	0.012 0.009 5.19E-06	0.003 2.50E-04 8.23E-05	5
Supercontig_454_5648	gi 166851824	cuticular protein analogous to peritrophins 3-C5 isoform 2 precursor [<i>Tribolium castaneum</i>]	81,8	411	1,00E-140	D/P	-3.010	-3.832	0.012	0.022	2
Contig6804	gi 662225759	PREDICTED: larval cuticle protein 1-like, partial [<i>Diaphorina citri</i>]	77,3	130	6,00E-32	D/P	-3.958	-4.807	2.23E-04	0.003	14
Contig11778	gi 389611489	cuticular protein PxutCPR146, partial [<i>Papilio xuthus</i>]	37,7	81,6	9,00E-14	D/P	-2.873	-3.795	0.016	0.018	19
comp173496_c1	gi 357614984	cuticular protein RR-2 motif 79 [<i>Danaus plexippus</i>]	85,7	139	2,00E-31	D/P	-2.432	-3.413	0.0349	0.047	9
Contig574	gi 646704638	Endocuticle structural glycoprotein SgAbd-2, partial [<i>Zootermopsis nevadensis</i>]	41,8	123	9,00E-31	D/P	-2.858	-3.754	0.016	0.014	10
comp171408_c0	gi 759035757	PREDICTED: pupal cuticle protein 20-like [<i>Cerapachys biroi</i>]	51,2	99	2,00E-20	D/P	-2.631	-5.491	0.039	0.020	9
Contig12583	gi 662197351	PREDICTED: endocuticle structural glycoprotein SgAbd-2-like [<i>Diaphorina citri</i>]	65,7	130	8,00E-33						4
comp107934_c0	gi 290558788	cuticular protein RR-1 motif 33 precursor [<i>Bombyx mori</i>]	65.15	157	2,00E-43						4
comp158087_c0	gi 665800845	PREDICTED: adult-specific cuticular protein ACP-22-like [<i>Microplitis demolitor</i>]	68.75	108	2,00E-24						20 (P<0.9)
comp170397_c0	gi 290560754	cuticular protein RR-2 motif 73 precursor [<i>Bombyx mori</i>]	89.19	75.9	2,00E-12						15
comp178445_c0	gi 662184166	PREDICTED: endocuticle structural glycoprotein SgAbd-9-like [<i>Diaphorina citri</i>]	53.92	101	2,00E-22						9
comp179689_c0	gi 157126722	cuticle protein, putative [<i>Aedes aegypti</i>]	75.36	123	6,00E-31	M/F	3.165	4.087	7.85E-06	3.64E-05	5

Contig1370	gi 501291136	cuticle protein, putative [Riptortus pedestris]	84.72	130	8,00E-33						9
Contig185	gi 646714048	Larval cuticle protein 2 [Zootermopsis nevadensis]	65.69	138	3,00E-34						9
Contig2710	gi 157129506	pupal cuticle protein, putative [Aedes aegypti]	75.29	129	1,00E-33	D/P	-2.827	-3.821	0.016	0.007	8
Contig6354	gi 501293980	cuticle protein, putative [Riptortus pedestris]	47.02	70.5	8,00E-10						4
Contig6573	gi 350406207	PREDICTED: endocuticle structural glycoprotein SgAbd-2-like [Bombus impatiens]	65.79	144	4,00E-37						9
Contig7304	gi 253735765	cuticle protein precursor [Acyrtosiphon pisum]	68.85	97.1	1,00E-21						9 (P<0.9)
Contig7887	gi 646704638	Endocuticle structural glycoprotein SgAbd-2, partial [Zootermopsis nevadensis]	54.81	127	3,00E-31						18
mira_454_rep_c244	gi 642924550	PREDICTED: endocuticle structural glycoprotein SgAbd-8 isoform X2 [Tribolium castaneum]	70.69	147	1,00E-38						filtered
Supercontig_454_1887	gi 642913097	PREDICTED: LOW QUALITY PROTEIN: larval cuticle protein A2B-like [Tribolium castaneum]	63.25	123	6,00E-31	D/P M/F	-2.545 2.983	-3.756 3.801	0.033 5.80E-05	0.011 2.26E-04	2
Supercontig_454_3182	gi 1706197	RecName: Full=Pupal cuticle protein C1B; AltName: Full=TM-PCP C1B; Short=TM-C1B [Tenebrio molitor]	78.18	87.4	1,00E-17	M/F	3.449	4.810	3.41E-06	6.92E-06	1
Supercontig_454_33795	gi 253735765	cuticle protein precursor [Acyrtosiphon pisum]	68.85	96.3	6,00E-21						3 (P<0.9)
Supercontig_454_5996	gi 662197351	PREDICTED: endocuticle structural glycoprotein SgAbd-2-like [Diaphorina citri]	71.91	124	3,00E-31						filtered
Supercontig_454_6096	gi 568250246	pupal cuticle protein 78E [Anopheles darlingi]	57.14	125	1,00E-31	D/P	-2.746	-4.719	0.029	0.017	9

Supplementary Table 2: Blast annotation of takeout, cytochrome P450, glutathione S-transferase and cuticular proteins and information about differential expression status.

Results of Blast: For all contigs that matched with a protein of interest in the non-redundant protein database, we indicate the protein ID (Blast NR ID) and its denomination (Annotation). For each match, we provide the percentage of identity (%ident), the score and the e-value. For each contig, we indicate in which comparison both DESeq2 and EdgeR found it to be found differentially expressed. D = domiciliary, P = peridomiciliary and S = sylvatic; M = male and F = female. The log2fold and the adjusted p-value for both packages are provided, as well as the cluster number allocated by HTScluster.

*Partie II : Analyse des transcriptomes
chimiosensoriels chez différentes espèces
du genre Rhodnius*

Chapitre 1 : Elaboration et annotation du transcriptome de *Rhodnius prolixus*

Résumé en français

Introduction : Comme nous l'avons vu dans la précédente partie avec le modèle *Triatoma brasiliensis*, les études comparatives de transcriptomes peuvent permettre la mise en évidence des différences d'expression de gènes selon l'environnement. Cependant, ces analyses comparatives nécessitent l'alignement des lectures obtenues par séquençage à haut débit des transcriptomes, sur un transcriptome de référence. Si l'espèce étudiée ne possède pas de référence, un assemblage *de novo* devra alors être effectué. La qualité de la référence doit être assez bonne pour capturer le catalogue de transcrits le plus complet possible. Si un génome de référence est disponible, on est à même de se demander si l'assemblage d'un transcriptome basé sur un génome de référence donnera de meilleurs résultats qu'un assemblage *de novo*. Le génome de *Rhodnius prolixus* ayant été récemment séquencé, cet insecte hématophage constitue un bon modèle pour comparer les méthodes d'assemblage de données transcriptomiques selon une procédure *de novo* ou basée sur un génome de référence.

Matériels et Méthodes : De la même manière que pour l'élaboration du transcriptome de référence de *T. brasiliensis*, différentes stratégies d'assemblage ont été testées à partir de données transcriptomiques provenant de transcriptomes chimiosensoriels de *R. prolixus* séquencés en 454, en Illumina paired-end ou en combinant les deux types de séquençage. Les mêmes logiciels et workflows d'assemblages *de novo* que ceux utilisés pour *T. brasiliensis* ont été comparés. Par ailleurs, l'alignement des lectures sur le génome de *R. prolixus* par Tophat suivi de l'assemblage du transcriptome par Cufflinks ont été réalisés, en fournissant ou non les indications d'annotation de ce génome. L'ensemble de ces tests a été comparé en suivant les mêmes critères que ceux utilisés chez *T. brasiliensis* (N50, nombre de contigs, complétude évaluée par CEGMA, proportion de contigs potentiellement chimériques). Par ailleurs, les contigs codant les protéines de liaison des molécules odorantes (OBPs et CSPs) ont été annotés par blast et utilisés pour donner une seconde évaluation de la complétude du transcriptome.

Résultats : De même que pour *T. brasiliensis*, parmi les assemblages *de novo* testés, l'assemblage hybride 454/Illumina a donné le meilleur résultat. Cependant, malgré le peu de recul existant sur le génome de *R. prolixus*, les assemblages du transcriptome basés sur cette référence montrent clairement de meilleurs résultats que la stratégie *de novo*. Contrairement au *de novo*, l'apport du 454 n'améliore pas l'assemblage pour cette technique, en revanche l'utilisation de plusieurs échantillons augmente la qualité du transcriptome de référence. Par ailleurs, l'apport des informations d'annotation du génome améliore l'assemblage. Toutefois, la combinaison des deux stratégies assemblage *de novo* et assemblage basé sur la référence a fourni le meilleur résultat et a notamment permis d'assembler des transcrits fractionnés issus des autres méthodes.

Conclusion : La combinaison des deux méthodes *de novo* et basée sur la référence génomique semble être une stratégie judicieuse pour construire un transcriptome de référence d'autant plus pour un génome nouvellement séquencé qui a fait l'objet de peu d'études. Par ailleurs, cette méthode a permis de détecter de nombreux transcrits non prédits dans le génome ainsi que de nombreux isoformes, ou encore des transcrits à cheval sur deux scaffolds du génome, informations qui pourraient être utiles pour l'amélioration de l'annotation du génome, ou de son assemblage par indication de scaffolds à fusionner.

Article 3: “Comparing de novo and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug *Rhodnius prolixus*” (Article in press in *Insect Biochemistry and Molecular Biology* 2015)

A. Marchant^{*1}, F. Mougell^{1, 2}, V. Mendonça^{1,3}, M. Quartier^{1,4}, E. Jacquin-Joly⁵, A. da Rosa³, E. Petit^{1,2}, M. Harry^{*1, 2}

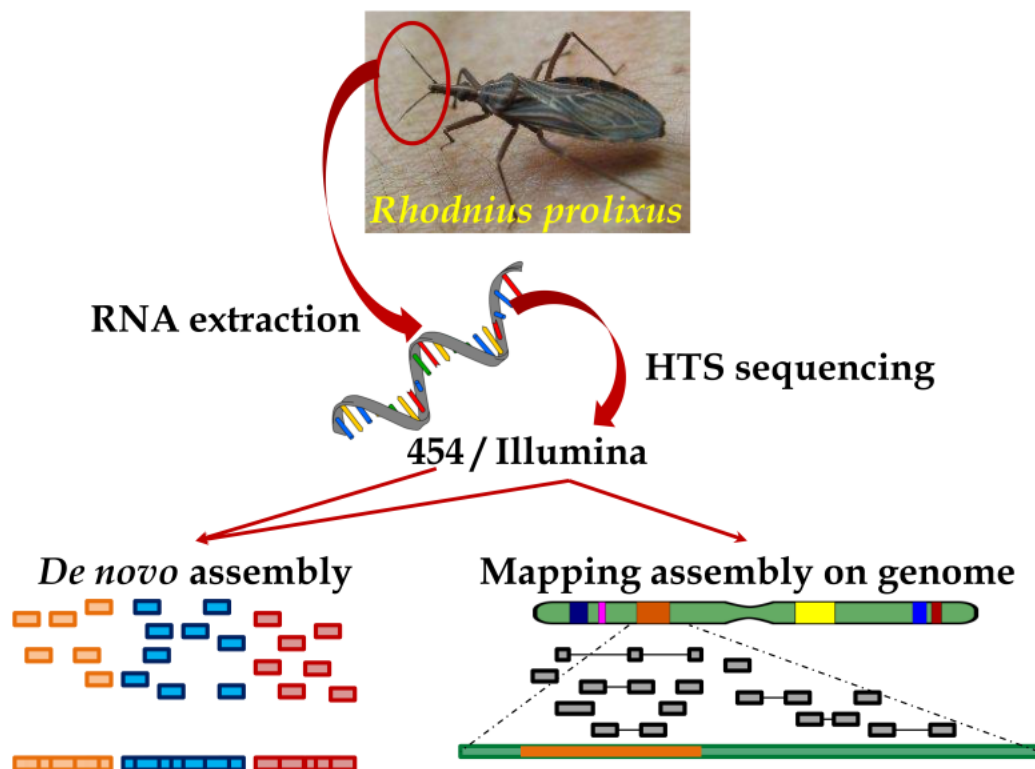
(1) Laboratoire Evolution, Génomes et Spéciation LEGS UPR 9034 - CNRS / UR 072 IRD, avenue de la Terrasse, Bâtiment 13, BP1 - 91198 Gif-sur-Yvette - France

(2) Université Paris Sud - Orsay - France

(3) Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP), Araraquara, Brazil

(4) University of Neuchâtel, Institute of Biology, Neuchâtel, Switzerland

(5) INRA, UMR 1392, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris), Versailles, France



Assembly	N50	Completeness	Chimeric contigs
<i>De novo</i> 454	577 bp	44.76%	0.41%
<i>De novo</i> Illumina	838 bp	91.13%	0.11%
<i>De novo</i> 454 + Illumina	1,217 bp	90.73%	0.73%
Mapping Cufflinks	1.656 bp	99.19%	0%
Mapping Cuffmerge	2,941 bp	98.79%	0%
<i>De novo</i> + Mapping Cuffmerge	3.022 bp	98.79%	0%

Abstract

High Throughput Sequencing capabilities have made the process of assembling a transcriptome easier, whether or not there is a reference genome. But the quality of a transcriptome assembly must be good enough to capture the most comprehensive catalog of transcripts and their variations, and to carry out further experiments on transcriptomics. There is currently no consensus on which of the many sequencing technologies and assembly tools are the most effective.

Many non-model organisms lack a reference genome to guide the transcriptome assembly. One question, therefore, is whether or not a reference-based genome assembly gives better results than *de novo* assembly. The blood-sucking insect *Rhodnius prolixus*—a vector for Chagas disease—has a reference genome. It is therefore a good model on which to compare reference-based and *de novo* transcriptome assemblies.

In this study, we compared *de novo* and reference-based genome assembly strategies using three datasets (454, Illumina, 454 combined with Illumina) and various assembly software. We developed criteria to compare the resulting assemblies: the size distribution and number of transcripts, the proportion of potentially chimeric transcripts, how complete the assembly was (completeness evaluated both through CEGMA software and *R. prolixus* proteome fraction retrieved). Moreover, we looked for the presence of two chemosensory gene families (Odorant-Binding Proteins and Chemosensory Proteins) to validate the assembly quality. The reference-based assemblies after genome annotation were clearly better than those generated using *de novo* strategies alone. Reference-based strategies revealed new transcripts, including new isoforms. However, a combination of both *de novo* and reference-based strategies gave the best result, and allowed us to assemble fragmented transcripts.

Keywords: *Rhodnius prolixus*, Chagas disease vectors, reference transcriptome, *de novo* assembly, referenced-based assembly, OBP, CSP

Introduction

Blood-sucking insects in the sub-family Triatominae (Hemiptera, Reduviidae) are vectors of *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae), which causes Chagas disease. Triatomines are obligate hematophages at all larval stages. *R. prolixus* has been a model for insect physiology and biochemistry for many decades (beginning with the early study by Wigglesworth, 1936). Because it was the main vector of Chagas disease in Central America, vector control programs have been developed to target domestic populations. Strong evidence now indicates that *R. prolixus* no longer transmits *Trypanosoma cruzi* in some northern countries of Latin America (Coura and Coura, 2013; Hashimoto and Schofield, 2012). But new vectors are now emerging that bring concerns about the potential health impacts.

Besides *R. prolixus*, some other *Rhodnius* species are now domiciliated—they are colonizing human dwellings. These include *Rhodnius ecuadoriensis* in Ecuador and *Rhodnius pallescens* in Panama. Species like *Rhodnius nasutus* and *Rhodnius neglectus* are in the process of domiciliation in Brazil. A

better understanding of the ecology, biochemistry and physiology of Triatomines will be helpful for developing new tools to control these disease vectors by disrupting chemical communication. In particular, we need to understand what is happening at the molecular level, by targeting chemosensory and behavioral genes that allow insects to interact with their environment (chemosensory recognition of food sources, breeding and oviposition sites) and with their congeners. Our goal here was to facilitate that understanding by building a reference transcriptome of *R. prolixus*.

Recent studies using High Throughput Sequencing (HTS) technologies have revolutionized the field of functional genomics and gene expression, and significantly improved genome annotation (see references in Carver et al., 2012). They opened the field to a wide range of applications from *de novo* sequencing to functional genomics and transcriptomics. Transcriptomics can be used to provide sequence information for *de novo* transcript assembly to improve gene annotations in species where the genome is not sequenced; detect differential expression at cell, tissue or population levels; detect new coding transcripts, non-coding RNAs or alternative splicing isoforms; and discover structural variations in transcripts such as SNPs or indels (see reviews: Burgess, 2013; Marguerat and Bähler, 2010; Mortazavi et al., 2008; Oshlack et al., 2010). Different sequencing technologies (Illumina, SOLiD, 454) and protocols (single end or paired ends) are currently available to perform RNA-Seq experiments. Several articles and reviews describe the current usage and limitations of these biological techniques (Wang et al., 2009; Wilhelm et al., 2010). Though many new algorithms and tools have been developed to efficiently process and analyze transcriptomic data (see for review Carver et al., 2012), there is no clear consensus on the best ways to perform assembly analysis.

To date, transcriptome studies on Triatominae have been mostly targeted sialotranscriptomes of various species because they are of pharmacological interest. Sialomes have been sequenced from salivary glands in species within two genera: *Rhodnius*, including *R. prolixus*, (Bussacos et al., 2011; Ribeiro et al., 2004), *R. brethesi* and *R. robustus* (Bussacos et al., 2011), and *Triatoma*, including *T. brasiliensis* (Santos et al., 2007) and *T. infestans* (Assumpção et al., 2008). The gene expression profile of *R. prolixus* ovarian follicle tissue have also been surveyed (Medeiros et al., 2011). Moreover, recently, Ribeiro et al., (2014) analyzed the transcriptome of the *R. prolixus* digestive tract using 454 technology. With the aim to build a reference transcriptome for the non-model insect *R. prolixus*, which lacked a reference genome, (Marchant et al., 2015a) performed *de novo* hybrid assembly using several programs and merged contigs from Illumina paired-end reads and 454 contigs corrected with Illumina reads.

A genome assembly for *R. prolixus* is available (http://metazoa.ensembl.org/Rhodnius_prolixus/Info/Index). In this study, we compare both *de novo* and reference-based strategies using various criteria, including contig lengths, N50, contig number, chimeric contigs, completeness and the fraction of the *R. prolixus* proteome retrieved.

Moreover, to validate the assembly quality, we evaluate the presence of a gene family; namely, Odorant-Binding Proteins (OBPs) and Chemosensory Proteins (CSPs).

Materials and Methods

Insects mRNA extraction and next-generation sequencing

We used two *R. prolixus* strains: the 077 strain maintained at the Insetário de Triatominae da Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP) - Araraquara and Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos/FIOCRUZ/Rio de Janeiro (males: PMEF; females: PFEF) and a strain maintained at the Institute of Biology, University of Neuchâtel, Switzerland (males: PMES; females: PFES). Samples included several individuals pooled by origin and by sex (Table 1). To target genes involved in the chemosensory system, RNA was extracted from the antennae and rostrum of Brazilian strains (PMEF and PFEF), and from the antennae, rostrum and head of Swiss strains (PMES and PFES) using the TRIzol® Reagent kit. Because sequencing with 454 technology requires a lot of RNA (5 to 10 µg), we pooled RNA aliquots extracted from the antennae and rostrum from UNESP laboratory strain (3 µg for PMEF, 3 µg for PFEF). To complete the sample, we used RNA extracted from the head, antennae and rostrum of the University of Neuchâtel strain. Since our goal was to favor chemosensory genes, we used less RNA from the latter strain (1 µg for PMES, 1 µg for PFES). The RNA mix (8 µg) was used as template for constructing a concatamerized, normalized cDNA library prepared and sequenced using 454 Roche GS FLX Titanium (½ Pico Titer Plate) by LGC Genomics GmbH (Berlin, Germany). We obtained 508,191 reads with a median length of 300 bp (279 bp after reads cleaning). Independent Illumina libraries, which requires less RNA, were prepared for three RNA samples (PMEF, PFEF, PMES) using 0.5 µg RNA aliquot per sample. Following rRNA removal (Ribo-Zero kit, Epicentre), the RNA was used to prepare cDNA libraries using the ScriptSeq™ mRNA-Seq Library Preparation Kit (Epicentre). Libraries were sequenced in paired-end reads (2x100 bp) on Illumina® HiSeq 2000 platform (Genomic Platform, IBENS, Institut de Biologie de l'École Normale Supérieure, Paris, France).

Sample	Population	Sexes	Number of individuals	Organs	Sequencing	Number of initial reads	Length median	Number of post-cleaned process reads
454_Rpro	Fiocruz+Suisse	M + F	28M+28F	Ant + ro + head	454	508,191	300	409,129
PMEF	Fiocruz	M	18	Ant + ro	Illumina paired-end	31,125,782 paired	101	29,670,809 paired
PFEF	Fiocruz	F	18	Ant + ro	Illumina paired-end	20,372,071 paired	101	18,986,527 paired
PMES	Suisse	M	10	Ant + ro + head	Illumina paired-end	42,792,624 paired	101	40,722,410 paired

Table 1: *R. prolixus* samples and associated sequencing information

The first letter of each Illumina sample name indicates the species (*prolixus*); the second letter indicates the gender (M for male and F for female) and the third letter is an insect rearing index (F for Fiocruz and S for Switzerland. With (ant): antennae, (ro): rostrum

Assembly of transcriptomes

- De novo assemblies

Three assemblies were compared using 454 reads alone, Illumina reads alone or mixing both 454 and Illumina reads for each library (Figure 1). Illumina and 454 reads were first cleaned using cleaned-1 criterion outlined in (Marchant et al., 2015a) with the Prinseq version 0.20.4 (Schmieder and Edwards, 2011). This version cleans forward and reverse reads simultaneously. When one read from a pair was discarded, the other was automatically eliminated. 454 reads were assembled with Newbler, CAP3 (Huang and Madan, 1999) and MIRA (Chevreux et al., 2004) according to the workflow described in detail in Marchant et al. (2015a, see assembly 4). Illumina reads were assembled using Trinity (Grabherr et al., 2011). A hybrid assembly was then performed according to Marchant et al. (2015a, see assembly 10). Possible homopolymer tracts were corrected by aligning Illumina reads on the 454 contigs using BWA (Li and Durbin, 2009). 454 corrected contigs and Illumina contigs were reassembled in “super-contigs” using CAP3 and repetitions were deleted using cd-hit-est (Li and Godzik, 2006).

- Reference-based assemblies

Two assemblies were generated for each of the three Illumina samples by mapping raw data on the genome assembly *Rhodnius prolixus* RproC1.25.dna.genome.fa (<http://metazoa.ensembl.org>) using TopHat (defaults options; Trapnell et al., 2009) and Cufflinks (Trapnell et al., 2010, Figure 1). The first assembly was done without specifying gene annotation and was named Cufflinks WA (without annotation; Figure 1). In the second assembly, we added the *R. prolixus* gene annotation (version *Rhodnius prolixus*.RproC1.25.gff3 of Ensembl) to Tophat (-G option) and Cufflinks (-GTF guide option) to guide the RABT (reference annotation based transcript) assembly. The output of RABT assembly included all annotated transcripts (with or without mapped reads), new isoforms and transcripts from novel loci (not initially annotated in the genome). Finally, we used Cuffmerge to merge the three reference-based assemblies from the three Illumina samples (PFEF, PMEF, and PMES) based on the genome annotation. For each contig, Cuffmerge output presented a class-code indicating whether the contig is a known transcript, a new isoform of a know transcript, or new transcripts. We also mapped 454 reads to the *R. prolixus* genome using GS Reference Mapper (www.454.com) with command line runMapping and option -cdna. Mapped reads were assembled with Cufflinks following reference genome annotation (RABT). Finally, we merged reference-based assemblies from 454 and Illumina reads with Cuffmerge to generate a hybrid 454/Illumina referenced-based transcriptome.

- De novo / Reference-based assembly

The existing *R. prolixus* assembly is fragmented (27,870 scaffolds). Consequently, transcriptome assembly based only on reference annotation may miss expressed genes that fall in the gaps of the genome. We thus decided to combine both *de novo* and reference-based assemblies. We used CAP3 with default options to reassemble 454 reads, Trinity contigs from all samples and

contigs from Cuffmerge in super-contigs (Figure 1). Contigs defined from Cuffmerge that were not assembled with CAP3 (“contigs singlets” in Figure 1) were added to the final assembly and duplicates were deleted using cd-hit-est.

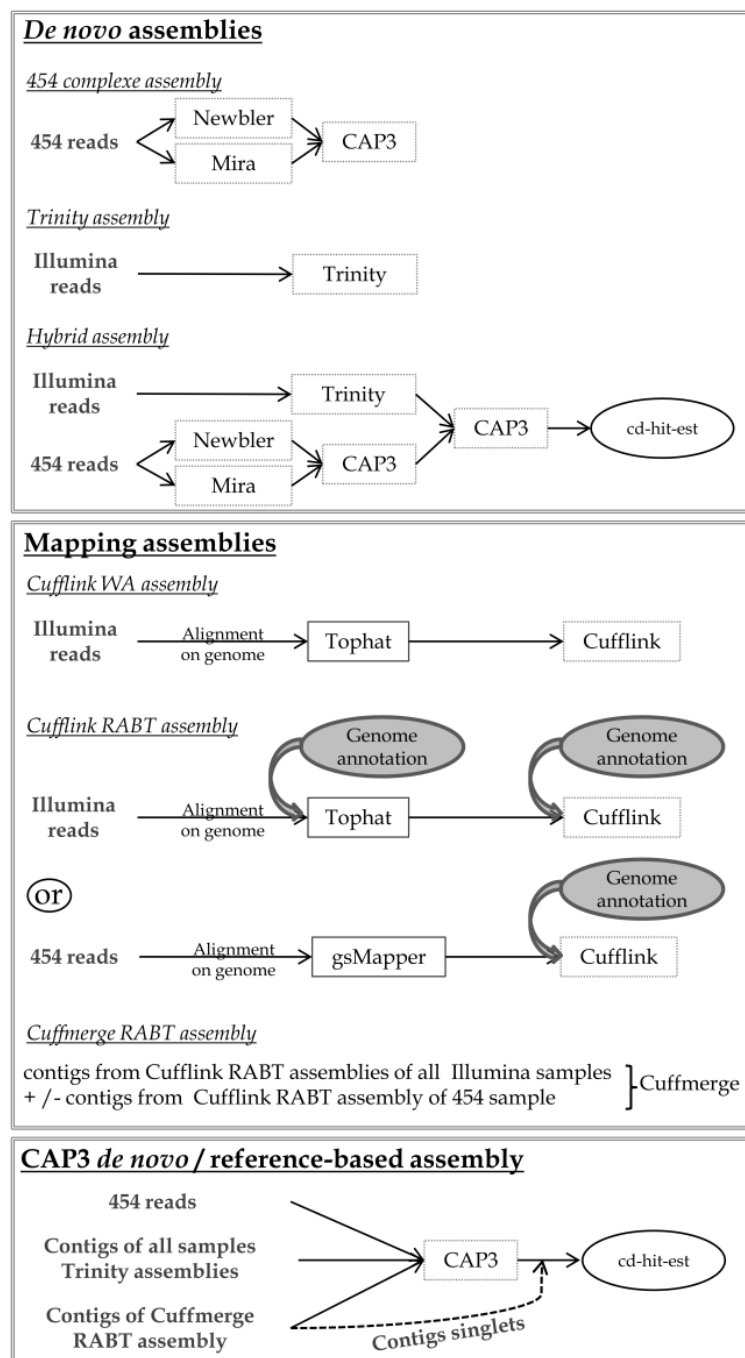


Figure 1: Workflow diagram of the different assemblies.

We used three different *de novo* assemblies Newbler, Mira and CAP3 were used to assemble 454 data (454 complex assembly), Trinity to assemble Illumina data (Trinity assembly), and CaP3 was used to merged the two previous assemblies (hybrid assembly). Two referenced-based assemblies were performed using for Illumina data the aligner software Tophat and for 454 data, GSMapper and the assembler Cufflink. In the “Cufflink WA assembly”, no annotation was given neither to aligner nor to assembler while in the “Cufflink RABT assembly”, a gtf file from the genome annotation was given to Cufflink. A joint *de novo* / reference-based assembly was built with CAP3 merging previous assemblies.

Criteria for determining assembly quality

We used several criteria to evaluate and compare the quality of our different assemblies, including contig lengths, N50, total number of bases assembled, and contig number. To check for the presence of chimeric contigs, we used a reciprocal Blast from the predicted proteome of *R. prolixus*, following the script described in Marchant et al., (2015) using an e-value of 10^{-6} . To estimate the completeness of the transcriptome, we used the CEGMA pipeline (Core Eukaryotic Genes Mapping Approach) that searches for sequences belonging to a set of ultra-conserved eukaryotic proteins (Parra et al., 2007). We also assessed completeness by looking at the fraction of *R. prolixus* proteome found in each assembly using TBLASTN (Madden, 2003). We used the best hit to estimate the fraction of proteome retrieved. Because reference-based assemblies include annotated transcripts, even if no reads were mapped, we excluded transcripts with no read mapped from these assemblies using BWA. We made two categories: partial alignment (more than 50% of the sequence) and complete alignment (more than 90%) using an identity index higher than 0.95. Finally, for hybrid *de novo* assemblies, we assessed completeness by mapping the transcripts from the different assemblies to the *R. prolixus* genome with GS Reference Mapper using `-cdna` option.

For OBP and CSP annotation, we searched genes within the 454 assembly and PMES Illumina reference-based assembly 15 using TBLASTN and available insect OBP and CSP amino acid sequences (*Drosophila melanogaster*, *Anopheles gambiae*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Pediculus humanus*, *Acyrtosyphon pisum*, *Locusta migratoria*) retrieved from GenBank and from Vieira and Rozas (2011) as queries. Candidate sequences were compared back to NCBI NR database using the BLASTX application to confirm annotation. We searched OBPs and CSPs for the presence of a signal peptide using SignalP 4.0 (Petersen et al., 2011) and predicted secondary structures using the Psipred server (Buchan et al., 2010). This search led to the description of OBP and CSP reference repertoires, respectively labeled from 1 to 27 and 1 to 19 (see the main paper concerning OPB, CSP section) and made it possible to describe a new OBP (OBP28). To detect isoforms, OBP and CSP transcripts were also searched by local BLASTX using this OBP and CSP reference repertoires as a database and the different transcriptome assemblies as queries. Hits with an e-value below 10^{-30} were counted as present.

Detection of contigs overlapping two or three scaffolds

Reference-based transcriptome assembly led to the discovery of numerous novel loci. We found some of them mapping at one or both extremities of genome scaffolds (1386 transcripts), questioning whether they were described completely. We thus compared these possibly truncated transcripts (PTT) with contigs generated from joint *de novo* and reference-based assembly, computing BLASTN with e-value threshold of e^{-60} and selecting only the first hit. Among the hits, we selected contigs aligned with the entire length of the PTT, except 20 bp on each side, and contigs recognized by a single PTT or by non-overlapping PTT. We extracted upstream and/or

downstream sequences of the contig matching PTT and searched their scaffold location with BLASTN against genome assembly (e-value = e^{-60} , one hit per query). Potential scaffold concatenation is proposed when i) alignment with the scaffold involved the entire upstream/downstream sequence extracted (except the 20 bp extremities), ii) upstream/downstream sequence and PTT matched different scaffolds and iii) the unaligned region of the scaffold that purportedly occurs in the pre-mRNA does not exceed 6000 bp (mean pre-mRNA size estimated from the predicted reference transcripts).

Results

De novo assemblies

To optimize the assembly, we tested three *de novo* assembly workflows: 454, Illumina, and hybrid Illumina/454 assemblies. After the cleaning process, between 80% (454) and 93-95% (Illumina) of the initial reads were retained, representing 409,129 reads for the 454 library and nearly 86 million paired-end reads for the three Illumina libraries (Table 1). Assemblies varied in number of contigs (Table 2). Assemblies built only with Illumina data have hundreds of thousands of contigs (PMEF: 557,578; PFEF: 234,677; PMES: 263,122). The assembly built only with 454 reads had 3.5–8.5 times fewer contigs (64,966). Hybrid assemblies with both 454 and Illumina contigs led to fewer contigs (assembly 5: 139,211; assembly 9: 62,169; assembly 13: 74,755). For all samples, the length distribution of the contigs show that the N50 is significantly better when the 454 and Illumina data are used in a hybrid assembly (assembly 5: 1,033; assembly 9: 810; assembly 13: 1,217 bp) compared to Illumina or 454 data alone (e.g., assembly 4: 627; assembly 8: 501; assembly 12: 838; assembly 2: 577 bp).

We used the Core Eukaryotic Genes Mapping Approach (CEGMA) to assess completeness of the assemblies. Both partial and complete CEGMA indices were generally low for assemblies that were based on a single data source (454 or Illumina), with the notable exception of the Trinity assembly generated with PMES Illumina reads (partial CEGMA reached 91.13% for this assembly). In this sample, RNA was extracted from the head, in addition to antennae and rostrum, which could explain why it was more complete. However, the complete CEGMA index was far lower than the partial one, even for PMES, indicating that the assembly was fragmented when using a single data source. CEGMA indices increased when we built hybrid assemblies that merged 454 and Illumina, except the complete CEGMA index for assembly 13. This may have been because of the assembly design, which only retained Trinity singlet-transcripts that were longer than 700 bp. Consequently, small transcripts detected as partial CEGMA proteins in the Trinity assembly were discarded in the hybrid assembly, decreasing the percentage of partial CEGMA set proteins. The fraction of the *R. prolixus* proteome retrieved in *de novo* assemblies agreed with CEGMA results: both complete and partial indices increased when 454 and Illumina were pooled for PFEF and PMEF samples (compare assembly 5 to 4 and 9 to 8 in Table 2).

As for CEGMA results, and likely for the same reason, the PMES sample followed the opposite trend: the fraction of partial protein sequences found decreased in hybrid assembly 13 compared to trinity assembly 12. Another notable observation was that both the CEGMA and the fraction of *R. prolixus* proteins was high for the PMES sample (assemblies 12 and 13) compared to PFEF (assemblies 8 and 9) and PMEF (assemblies 4 and 5). Since the PMES RNA sample was enriched with head tissue, the whole transcriptome had higher representation with every workflow used. The percentage of potential chimeric contigs was low in all assemblies (highest is 0.729% in assembly 13) and slightly increased in hybrid assemblies compared to assemblies obtained from a single data source. The procedure for mapping the assembled contigs to the reference *R. prolixus* genome could provide additional information on the reliability of transcriptome assemblies. By mapping hybrid *de novo* assembly contigs to the *R. prolixus* genome using GS reference Mapper, we obtained 98.59% mapped contigs for assembly 5, 97.93% for assembly 9, and 98.09% for assembly 13.

Reference-based assemblies

The reference-based assembly method was more efficient than the *de novo* one, whatever the data source: Illumina PMEF, PFEF and PMES or 454. The contig length increased in all cases and the percentage of chimeric contig was similar or lower. Except for assembly 10, CEGMA indices sometimes increased dramatically. The improvement was particularly striking when genome annotation was taken into account: partial CEGMA index was above 95% in all cases, the percentage of potential chimeric was null, and the N50 reached 2000 bp for some samples.

Finally, the best reference-based assemblies obtained were those that merged different assemblies from male and female samples: the N50 rose sharply to about 3,000 bp and the CEGMA reached 98.79%. However, the addition of 454 data did not significantly improve the results—assemblies 16 and 17 performed similarly. We also evaluated the fraction of *R. prolixus* proteome using TBLASTN. In accordance with CEGMA results, the highest percentages were found when annotation was taken into account and different samples were merged (up to 93.5% in assemblies 16 and 17). This index was not expected to reach 100%, however, as the transcriptome was built with RNA extracted from specific tissues (antennae, rostrum and head for PMES) where some genes are probably not expressed.

Finally, reference-based assemblies made it possible to describe new isoforms and loci compared to the reference transcriptome. In assembly 17, we detected 21,355 new isoforms of the reference transcripts and 142,096 unknown or intergenic transcripts described on 137,361 loci. While, the cDNA library was depleted of ribosomal RNA, long RNA, miRNA and siRNA could probably make up an important part of these new loci.

De novo / Reference-based assemblies

As expected, CAP3 pooled contigs in super-contigs increased the N50 (3,022 bp) and the average length distribution (2,170 bp; Table 2). The percentage of complete CEGMA set protein was the highest (93.55%). However, the percentage of *R. prolixus* proteins retrieved did not follow the same way—both partial and complete indices were lower than those found in assemblies 16 and 17 (85.9 versus 90.30 or 90.4 versus 91.9 for the complete index). CAP3 super-contigs seem to have remodeled some of the transcripts and were consequently protein encoded.

Sample (assembly)	Contig number	total length	Length distribution (bp)			CEGMA (%)		Potential chimeric contigs proportion (%)	Fraction of <i>R. prolixus</i> proteome (%)	
			max	average	N50 (bp)	complete	partial		complete	partial
1) <i>Rhodnius prolixus</i> . RproC1.25.cdna.all.fa	15,725	17,270,277	34,028	810	1,503	84.68	96.37			
2) 454 complex assembly	64,966	32,955,859	4,045	507	577	25.81	44.76	0.411	5.2	10.5
3) 454 Cufflinks RABT	23,427	24,817,294	248,692	1,059	1,942	85.06	96.77	0	37.5	39.4
4) PMEF Trinity	557,578	293,961,617	11,597	527	627	8.06	30.65	0.018	4.5	15.3
5) PMEF + 454 hybrid <i>de novo</i>	139,211	137,194,131	11597	986	1,033	29.44	54.84	0.203	7.9	17.6
6) PMEF Cufflink WA	182,559	352,947,789	17,685	1,933	2,538	30.24	66.53	0.244	10.2	22.9
7) PMEF Cufflink RABT	209,914	373,294,695	248692	1,778	2,415	89.11	97.58	0	89.9	93.3
8) PFEF Trinity	234,677	105,076,048	5,386	448	501	2.02	9.68	0.036	2.9	7.4
9) PFEF + 454 hybrid <i>de novo</i>	62,169	50,327,189	5,386	810	885	24.60	37.90	0.457	6.7	12.5
10) PFEF Cufflink WA	128,328	112,738,425	7,257	879	1,063	5.24	19.76	0.134	4.1	8.9
11) PFEF Cufflink RABT	148,600	137,733,087	248,692	927	1,170	86.29	97.58	0	71.0	73.7
12) PMES Trinity	263,122	165,386,157	21,462	629	838	41.98	91.13	0.111	18.6	43.1
13) PMES + 454 hybrid <i>de novo</i>	74,755	84,977,419	21,462	1,137	1,217	61.29	90.73	0.729	19.5	38.5
14) PMES Cufflink WA	63,622	64,840,114	21,462	1,019	1,357	77.02	93.55	0.569	30.3	46.7
15) PMES Cufflink RABT	81,459	88,125,904	248,692	1,082	1,656	92.34	99.19	0	82.2	84.9
16) Assemblies 7 + 11 + 15 → Cuffmerge	192,429	389,834,701	248,692	2,026	2,995	92.74	98.79	0	90.3	93.5
17) Assemblies 3 + 7 + 11 + 15 → Cuffmerge	191,417	382,696,685	256,221	1,999	2,941	92.74	98.79	0	90.4	93.5
18) Assemblies 4 + 5 + 8 + 9 + 12 + 13 + 16 → CAP3 + singlets 20	165,102	358,268,586	256,221	2,170	3,022	93.55	98.79	0.193	85.9	91.9

Table 2 : Metrics used to evaluate the assemblies

Detection of contigs overlapping two or three scaffolds

From the 1,386 potentially truncated transcripts (PTT) produced by reference-based assembly, 1,167 matched specific contigs of the joint *de novo* / reference-based assembly, but only 602 of them were aligned over their entire length to 428 different contigs. Four contigs were aligned with two non-overlapping PTT. In one case, the two PTT were mapped on the same scaffold and considered as different transcripts in the reference-based assembly, while they were concatenated in a single transcript in the joint *de novo* / reference-based assembly. In the three other cases, PTT were identified on different scaffolds, suggesting potential physical linkage between scaffolds (Table 3). One hundred and twenty contigs were aligned to two or more overlapping PTT and discarded. Of the 303 contigs aligned to a single PTT, 73 aligned to a single scaffold making up the PTT, 38 led to ambiguous matches, while 119 aligned to two and 73 to three different scaffolds. Of these 192 contigs, 84 were considered valid when the size criterion was applied. The final set of 84 contigs suggests possible scaffold concatenation. It is listed in supplementary table 1.

OBP and CSP searches

By searching genes of the multigenic families OBP and CSP, we aimed to evaluate the completeness of our assemblies on a smaller scale. In 454 assembly 2, we identified 17 of the 27 OBP and 14 of the 19 CSP annotated from the *R. prolixus* genome (Table 3). In the hybrid *de novo* assembly 13, 21 OBPs and 12 CSPs were identified. In the reference-based assembly 15, the complete dataset was identified except OBP15, which is missing in all our assemblies regardless of the method used (Table 4). In reference-based assemblies, one additional OBP—not previously annotated in the *R. prolixus* genome and predicted as RPRC009906—was retrieved with a significant hit with the *Rhodnius prolixus* OBP2 described in Zhou et al., (2010a). After mapping Illumina reads on the *R. prolixus* genome, we found more exons than predicted in RPRC009906. The length of the transcript was increased and the translated sequence identified as an OBP (Table 5). Many of the CSP and OBP detected were observed through various isoforms, the number of which was highest when Illumina data was used and when assembly was performed using reference-based methods (overall: 59 putative OBP isoforms, 34 putative CSP isoforms for assembly 17 vs 19 putative OBP isoforms, 15 putative CSP isoform for assembly 2). Up to five isoforms were found per type of OBP, and only three per type of CSP. By contrast, fewer isoforms were described with assembly 18, where low divergent sequences were grouped in a single isoform.

	Scaffold 1			Coordinates Transcript Cuffmerge in Scaffold 1				Scaffold 2			Coordinates Transcript Cuffmerge in Scaffold 2			
Contig Cap3	Scaffold name	length	orientation /contig CAP3	Transcript Cuffmerge name	start	end	orientation	Scaffold 2 name	length	orientation / contig CAP3	Transcript Cuffmerge name	start	end	orientation
Contig41510	GL565382	1121	+	TCONS_00187737	1	1068	+	GL545587	2376	-	TCONS_00000523	1330	2376	+
Contig13674	GL567822	1157	+	TCONS_00188909	1	294	+	GL567822	1157	+	TCONS_00188910	542	1157	+
Contig41702	GL556040	1463	+	TCONS_00020998	1	1449	+	GL568106	1011	-	TCONS_00189058	1	703	+
Contig16237	GL545355	1422	-	TCONS_00000136	70	1422	+	GL545423	1226	-	TCONS_00000238	535	1226	+

Table 3: Results of the detection of contigs overlapping several scaffolds

Some Cuffmerge contigs mapping at extremities of different genome scaffolds were gathered together by CAP3, giving clues for scaffolds concatenation.

The CAP3 contigs which assemble cuffmerge contigs are shown here. We indicate scaffold names, their length and their orientation compared to CAP3 contigs. Cuffmerge transcript and their coordinates mapping in scaffold are also showed.

*Partie II : Analyse des transcriptomes chimiosensoriels chez
différentes espèces du genre Rhodnius*

VectorBase (Vb) prediction ID	identified as	2) 454 complex assembly	13) PMES + 454 hybrid <i>de novo</i>	16) PMES Cufflink RABT	17) Assemblies 3 + 7 + 11 + 15 → Cuffmerge	18) Assemblies 4 + 5 + 8 + 9 + 12 + 13 + 16 → CAP3 + singlets 20
RPRC010284-PA	OBP1	1	2	2	3	4
RPRC000378-PA	OBP2	0	0	1	1	1
RPRC000527-PA	OBP3	0	1	1	1	1
RPRC000274-PA	OBP4	0	2	3	4	3
no Vb predicted protein GL562839 735819-746677	OBP5	0	1	1	1	1
RPRC004413-PA	OBP6	1	1	1	3	2
RPRC000311-PA	OBP7	1	1	1	1	1
RPRC000564-PA	OBP8	0	0	2	2	2
RPRC000194-PA	OBP9	0	0	1	1	1
RPRC000560-PA	OBP10	2	1	2	2	2
RPRC004408-PA	OBP11	1	1	1	3	3
RPRC000492-PA	OBP12	1	0	1	1	1
RPRC000177-PA	OBP13	2	1	1	2	1
RPRC004407-PA	OBP14	1	2	1	2	2
RPRC008413-PA	OBP15	0	0	0	0	0
RPRC011610-PA	OBP16	0	3	1	2	1
RPRC000118-PA	OBP17	1	1	1	3	2
RPRC001924-PA	OBP18	1	1	1	1	1
RPRC000593-PA	OBP19	0	1	2	3	2
RPRC007008-PA	OBP20	1	2	2	4	4
RPRC001925-PA	OBP21	1	1	3	5	3
RPRC000115-PA	OBP22	1	1	3	4	2
RPRC000174-PA	OBP23	1	2	1	2	1
RPRC009905-PA	OBP24	1	2	1	2	2
RPRC000257-PA	OBP25	0	0	1	1	1
RPRC000182-PA	OBP26	1	1	3	3	3
RPRC009914-PA	OBP27	1	2	1	1	1
RPRC009906-PA corrected	OBP28	0	0	1	1	1
RPRC013223-PA	CSP1	1	1	1	2	1
RPRC013221-PA	CSP2	1	1	1	2	1
RPRC013280-PA	CSP3	1	1	1	2	1
RPRC013285-PA	CSP4	1	1	1	1	1
RPRC013218-PA	CSP5	1	2	1	3	2
RPRC000332-PA	CSP6	1	1	1	1	1
RPRC010107-PA	CSP7	0	1	2	2	2
RPRC000631-PA	CSP8	2	1	2	4	2
RPRC000470-PA	CSP9	1	0	1	2	1
RPRC007979-PA	CSP10	1	0	1	1	1
RPRC011935-PA	CSP11	1	1	1	1	1
RPRC011928-PA	CSP12	1	1	1	2	1
RPRC011926-PA	CSP13	1	0	1	2	1
RPRC006791-PA	CSP14	0	0	1	1	1
RPRC000546-PA	CSP15	0	0	1	1	1
RPRC000327-PA	CSP16	0	0	2	3	1
RPRC007134-PA	CSP17	1	1	1	2	2
RPRC000242-PA	CSP18	0	0	1	1	1
RPRC000126-PA	CSP19	1	1	1	1	1
Total OBP (total isoforms)	OBP	17 (19)	21 (30)	27 (40)	27 (59)	27 (49)
Total CSP (total isoforms)	CSP	16 (15)	12 (13)	19 (22)	19 (34)	19 (23)
Nb of contigs		34	43	62	93	72

Table 4: Number of putative isoforms by OBPs or CSPs contig matching according to the assembly (see Material and Method section). OBP and CSP transcripts were searched by local BLASTX with the OBP and CSP genes annotated from the *Rhodnius prolixus* genome as a database and the different transcriptome assemblies as queries. Only hits with an e-value below 10⁻³⁰ were counted as present. The OBP5 wasn't predicted in the genome. We identified it the scaffold GL562839 between the 735,819 and the 746,677 bases.

Transcript	NR	Annotation	%ident	Score	E value	Cysteins	P Signal	α Helice
RPRC009906-RA corrected	gi 240247219	odorant-binding protein RproOBP2 precursor, partial <i>Rhodnius prolixus</i>	34.19	61.6	2,00E-09	6	NO	7

Table 5: Feature of the new OBP annotated in reference-based assemblies. The transcript RPRC009906-RA was corrected taking into account read mapping information. Details about the best hit in NR database of this corrected transcript are given: the sequence NR ID; the annotation of NR sequence, the percent identity between NR sequence and our transcript, the blast Score and the e-value. The corrected transcript was also analyzed for specific features of OBP family: number of cysteins on a clustalW multiple alignment with other RproOBP, presence of a peptide signal (P Signal) and number of α Helices. Except, the peptide signal, all features confirm the OBP nature of this new transcript.

Discussion

Our work focused on comparing different transcriptome assembly processes, compiling various data sources and information (454 and Illumina reads, genome assembly and its predicted gene set). Indeed, the aim of this study was to perform an accurate and complete reference transcriptome to run further differential expression analyses for chemosensory genes and for downstream transcriptomic studies such as functional genomics or SNP discovery. Although many tools have been developed to assess genome assemblies (e.g. Bradnam et al., 2013; Clark et al., 2013; Gurevich et al., 2013), standards for systematically assessing the quality of transcriptome assemblies are still in development.

Transcriptome reconstruction can be classified into reference-based and *de novo* assembly approaches. *De novo* assembly approaches provide an important solution when the reference genome is not available, or is incomplete. Since there are a number of different tools and parameters that can be used to reconstruct transcripts, little is known about the impact of sequencing data on *de novo* assembly, even if there are metrics available to produce the highest quality transcriptome assemblies (O’Neil and Emrich, 2013).

Studies have used such metrics to compare different *de novo* approaches from varying transcriptome data sets, including data from extensive *in silico* computer simulations (Vijay et al., 2013) or biological data (Marchant et al., 2015a). In the present study, the results of the *R. prolixus de novo* transcriptome assembly were consistent with the results of Marchant et al. (2015a), where hybrid assembly of *de novo* 454 / Illumina increased contig size, reduced contig number and increased the completeness of the transcriptome compared to simple *de novo* assemblies based on only one technology (Illumina or 454 separately). Similar results have been generated for transcriptome analysis for other non-model organisms; for example, the Japanese quail (Finseth and Harrison, 2014), where the indel and frameshift errors inherent to the 454 transcriptome were corrected with Illumina reads. Martínez-Barnetche et al. (2012) made one of the first transcriptome hybrid 454 / Illumina assemblies and mapped it to the heterologous genome of *A. darlingi*. 92% of their transcripts were mapped and covered 90% of the core eukaryotic genome, confirming the reliability of the hybrid assembly method.

When a reference genome is available, reference-based approaches have been implicitly considered to be more effective than *de novo* assembly (Martin and Wang, 2011), but few studies have compared the two strategies (but see Hornett and Wheat, 2012). According to Martin and Wang (2011), the reference-based transcriptome assembly strategy has several advantages. In particular, reference-based transcriptome assembly has high sensitivity, making it possible to assemble transcripts of low abundance and to discover novel transcripts that are not present in the existing annotation. Reference-based assembly also has a few drawbacks; for example, it is difficult to assemble *trans*-spliced genes and it depends on the quality of the reference genome being used. Moreover, when a reference genome is not available for the organism concerned, it is sometimes possible to use the reference from a closely related species, even if transcripts from divergent genomic regions would be missed (e.g. Hornett and Wheat, 2012; Ward et al., 2012).

Since there is a reference genome available for *R. prolixus*, we were able to compare the two strategies. As expected, reference-based assembly provided better results, with an N50 and completeness (CEGMA and fraction of reference proteins) greater than those obtained from *de novo* assemblies. Among them, the RABT (reference annotation based transcript) assembly led to the highest N50 and the most complete transcriptome, with a reasonable number of transcripts. Moreover, this approach made it possible to find isoforms of known transcripts and to define novel loci in the genome transcribed to RNA. With both *de novo* and reference-based assemblies, using a sample from RNA extraction of a wider range of tissues (antenna, rostrum and head *versus* just the antenna and rostrum) led to the most complete transcriptome, demonstrated by the high CEGMA. Searching for OBPs and CSPs in our various assemblies confirmed that reference-based assemblies are more reliable than *de novo* ones. Nevertheless, combining *de novo* and RABT assemblies seems to be a powerful strategy when the reference genome is fragmented. It enabled us to increase the length of contigs, led to a more complete transcriptome and could be used to generate first clues to reassembling a fragmented reference genome scaffold.

With this study, we quantified how much an assembly could be improved when a reference genome is available, as for *R. prolixus*. Moreover, the transcriptome appears to be a powerful complement to the genome, making it possible to both validate predicted genes and detect exon shuffling that generates various isoforms from a single gene. Combined, genomic and transcriptomic data help to define new genes that are not predicted with the use of an automatic annotation, as demonstrated from the OBP and CSP gene family.

Acknowledgments

We would like to thank Patrick Guerin of the Institute of Biology at the University of Neuchâtel, Switzerland for the *R. prolixus* strain, Rachel Legendre and Claire Toffano of Institut de Génétique et Microbiologie CNRS - UMR 8621 who gave us the script for 454 contig correction. Our thanks go to Nicolas Pollet who helped us run GS reference Mapper. We are also very grateful to the engineers of the bioinformatics platforms Genouest at the University of Rennes 1 and platforms Migale at the INRA of Jouy en Josas. This study was funded by the French Agence Nationale de la Recherche (ADAPTANTHROP project, ANR-097-PEXT-009) and supported by LabEx BASC (University Paris Saclay, France). AM was funded by the Idex Paris Saclay, France.

Supplementary materials

Contig	GL1	Length GL1	Start1	End1	Orientation	GL2	GL2	Length GL2	Start2	TCONS	Orientation	GL3	Length GL3	Start3	End3	Orientation
Contig9042	-	-	-	-	-	GL563356	1246	771	1246	TCONS_00186872	-	GL545768	6651	746	2411	-
Contig41876	-	-	-	-	-	GL570774	1047	1	1029	TCONS_00190258	-	GL546617	5369	402	1605	+
Contig16439	-	-	-	-	-	GL568438	1060	1	769	TCONS_00189203	-	GL548047	2335	1311	1469	-
Contig41685	-	-	-	-	-	GL567857	1076	34	1076	TCONS_00188930	-	GL548654	26949	26309	26949	-
Contig41722	-	-	-	-	-	GL568488	1564	20	1564	TCONS_00189226	-	GL548979	3261	73	2390	-
Contig42032	-	-	-	-	-	GL571570	1186	1	1073	TCONS_00190637	-	GL551747	1911	8	262	-
Contig3057	-	-	-	-	-	GL567600	1021	1	1021	TCONS_00188775	-	GL552802	1329	242	1329	-
Contig41771	-	-	-	-	-	GL569152	1130	40	1130	TCONS_00189534	-	GL553984	4866	3548	4346	-
Contig41658	-	-	-	-	-	GL567432	1016	656	1016	TCONS_00188695	-	GL556721	5724	3692	5497	-
Contig8835	-	-	-	-	-	GL564835	1332	499	1332	TCONS_00187490	-	GL557665	25088	4863	5209	+
Contig16867	-	-	-	-	-	GL566853	1045	112	1045	TCONS_00188420	-	GL558123	1848	1040	1847	+
Contig3100	-	-	-	-	-	GL566519	1216	5	1216	TCONS_00188253	-	GL558367	5948	1902	2591	+
Contig41565	-	-	-	-	-	GL566237	1043	6	1043	TCONS_00188116	-	GL559653	1531	418	1386	-
Contig3109	-	-	-	-	-	GL570443	1066	1	1066	TCONS_00190096	-	GL562284	26584	3712	4193	+
Contig41750	-	-	-	-	-	GL568857	1120	3	1120	TCONS_00189399	-	GL563000	1254935	1	396	+
Contig41812	-	-	-	-	-	GL569663	1025	56	1025	TCONS_00189755	-	GL563050	710082	708586	709357	-
Contig39522	-	-	-	-	-	GL551932	1285	1	1280	TCONS_00013170	-	GL563169	205613	1309	3105	+
Contig41859	-	-	-	-	-	GL570519	1176	1	1164	TCONS_00190132	-	GL564257	1362	44	216	-
Contig16846	-	-	-	-	-	GL557855	4505	1	2432	TCONS_00024064	-	GL564793	1428	1245	1426	+
Contig41969	-	-	-	-	-	GL571972	1076	48	1076	TCONS_00190832	-	GL566218	1285	632	1282	+
Contig16979	-	-	-	-	-	GL560365	1659	1153	1659	TCONS_00027624	-	GL566969	1043	116	739	-
Contig41365	-	-	-	-	-	GL563205	1049	213	1049	TCONS_00186799	-	GL568110	1725	894	1393	-
Contig16762	-	-	-	-	-	GL556072	1982	1	1319	TCONS_00021076	-	GL570379	1250	424	1129	+
Contig16919	-	-	-	-	-	GL558982	2245	2	2245	TCONS_00025970	-	GL570924	1018	649	1015	+
Contig16727	-	-	-	-	-	GL555347	2020	18	2020	TCONS_00019965	-	GL571622	1079	1	1079	+
Contig16474	-	-	-	-	-	GL550064	2728	1	1389	TCONS_00009362	-	GL572005	1211	755	1206	+
Contig16952	-	-	-	-	-	GL566624	1625	286	1625	TCONS_00188302	-	GL572015	1692	918	1685	+
Contig41641	-	-	-	-	-	GL572598	1071	1	1057	TCONS_00191143	-	GL572676	1627	1001	1626	+
Contig16243	GL545519	1970	1087	1645	+	GL545329	2228	1	2207	TCONS_00000101	-	-	-	-	-	-
Contig9124	GL545813	3036	1887	2555	+	GL568073	1048	1	1043	TCONS_00189039	-	-	-	-	-	-
Contig16278	GL546208	1365	479	656	-	GL547921	1101	1	1101	TCONS_00005334	-	-	-	-	-	-
Contig16290	GL546417	34888	29551	30403	+	GL568822	1127	134	1127	TCONS_00189383	-	-	-	-	-	-
Contig8985	GL547030	75491	1098	1396	-	GL569695	1528	1	1515	TCONS_00189771	-	-	-	-	-	-
Contig16324	GL547140	1606	604	1239	-	GL570550	1334	1	618	TCONS_00190152	-	-	-	-	-	-

Contig16494	GL550598	30455	1197	1576	-	GL564751	1281	1	1281	TCONS_00187458	-	-	-	-	-	-
Contig41665	GL550834	1113	389	1111	+	GL566108	1074	1	1072	TCONS_00188057	-	-	-	-	-	-
Contig16885	GL551878	2049	46	538	-	GL560536	1081	1	990	TCONS_00027782	-	-	-	-	-	-
Contig6354	GL552237	1246	808	965	+	GL572360	1065	489	1065	TCONS_00191011	-	-	-	-	-	-
Contig16591	GL552546	3637	1977	2462	+	GL572481	1011	3	1011	TCONS_00191080	-	-	-	-	-	-
Contig2595	GL553908	1325	869	1026	+	GL569642	1149	19	1149	TCONS_00189746	-	-	-	-	-	-
Contig16663	GL554195	6926	2212	4767	-	GL568128	1003	3	1003	TCONS_00189065	-	-	-	-	-	-
Contig16704	GL554843	1610	982	1272	+	GL565738	1001	1	960	TCONS_00187878	-	-	-	-	-	-
Contig16857	GL558028	5063	3462	3988	+	GL572285	1388	4	1388	TCONS_00190975	-	-	-	-	-	-
Contig16927	GL559216	5715	1089	1222	+	GL566686	1050	1	1013	TCONS_00188333	-	-	-	-	-	-
Contig8861	GL561685	28701	3925	5965	-	GL553708	1711	95	1711	TCONS_00016938	-	-	-	-	-	-
Contig17197	GL561690	16317	14292	15684	+	GL564770	1003	1	715	TCONS_00187466	-	-	-	-	-	-
Contig17751	GL562518	184371	2240	3915	-	GL567584	1068	43	1068	TCONS_00188769	-	-	-	-	-	-
Contig16635	GL563210	1068	1	945	+	GL569850	1003	249	1003	TCONS_00189830	-	-	-	-	-	-
Contig41400	GL563778	1025	14	626	+	GL558176	1248	120	1248	TCONS_00024594	-	-	-	-	-	-
Contig1218	GL564541	1033	682	970	-	GL564956	1055	474	966	TCONS_00187549	0	-	-	-	-	-
Contig13452	GL564847	1063	1	640	+	GL558221	1313	1	1202	TCONS_00024657	-	-	-	-	-	-
Contig41543	GL565966	1171	602	1162	-	GL570936	1233	712	1233	TCONS_00190330	-	-	-	-	-	-
Contig3074	GL566178	1220	3	870	+	GL572119	1149	1	167	TCONS_00190896	-	-	-	-	-	-
Contig3141	GL568419	1056	92	1048	-	GL571306	1362	1	1336	TCONS_00190511	-	-	-	-	-	-
Contig13503	GL568686	1754	1005	1746	+	GL554486	1201	1	1006	TCONS_00018472	-	-	-	-	-	-
Contig41783	GL569395	1184	609	979	-	GL555579	1594	125	1594	TCONS_00020403	-	-	-	-	-	-
Contig41855	GL570375	1007	2	276	+	GL570376	1000	1	985	TCONS_00190071	-	-	-	-	-	-
Contig41930	GL571463	1326	5	156	+	GL547047	17989	6217	17989	TCONS_00003498	-	-	-	-	-	-
Contig1431	GL571479	1247	1	695	+	GL558727	2345	1	2034	TCONS_00025517	-	-	-	-	-	-
Contig41829	GL571801	1483	802	1130	-	GL571940	1210	1	1153	TCONS_00190820	-	-	-	-	-	-
Contig41994	GL572331	1123	50	522	+	GL550381	1328	1	1327	TCONS_00010039	-	-	-	-	-	-
Contig42030	GL572680	1168	4	420	+	GL559819	3584	1449	3584	TCONS_00027010	-	-	-	-	-	-
Contig14972	GL568271	1263	1	1263	-	GL568271	1263	1	1261	TCONS_00189122	-	GL545419	1503	351	1081	+
Contig2726	GL559157	1098	370	615	+	GL567554	1094	1	275	TCONS_00188755	-	GL547373	26484	3191	4278	+
Contig16349	GL547628	1165	45	652	+	GL568921	1371	1	1363	TCONS_00189424	-	GL547587	4778	4196	4395	+
Contig8974	GL564489	1376	906	1216	-	GL568336	1216	28	1216	TCONS_00189154	-	GL548512	6123	139	838	+
Contig16470	GL549994	10850	8625	9610	+	GL572568	1209	1	1209	TCONS_00191130	-	GL550620	1374	399	1255	+
Contig16395	GL548468	5382	71	1813	+	GL568852	1147	501	1147	TCONS_00189395	-	GL553827	6532	5271	6356	+
Contig16430	GL549293	5817	559	971	-	GL547838	1247	2	1247	TCONS_00005133	-	GL553909	2062	5	2062	-
Contig41845	GL554500	17319	33	1285	-	GL568728	1153	1	1143	TCONS_00189339	-	GL554572	2067	2	913	-
Contig25432	GL563014	716454	4197	8662	-	GL550812	2446	1946	2446	TCONS_00010916	-	GL555510	2008	728	1917	+
Contig14954	GL548149	1399	759	1399	-	GL545715	1423	47	1423	TCONS_00000802	-	GL558026	1907	58	1906	-

Contig8837	GL559358	2733	116	2374	-	GL550163	1322	1	1314	TCONS_00009616	-	GL559617	1480	11	474	-
Contig16504	GL550795	1472	7	384	+	GL566580	1407	1	769	TCONS_00188285	-	GL559739	2901	1022	2850	+
Contig5491	GL569548	1097	433	914	+	GL571887	1064	1	749	TCONS_00190792	-	GL560092	1293	193	1176	+
Contig16992	GL552307	1799	1229	1413	-	GL572669	1016	180	1016	TCONS_00191168	-	GL560490	1851	1048	1476	+
Contig7233	GL572829	1094	82	479	-	GL572828	1064	1	1061	TCONS_00191249	-	GL561497	46345	1	1734	+
Contig3026	GL555138	1677	1125	1370	-	GL569909	1012	65	1012	TCONS_00189858	-	GL564429	1143	208	1143	+
Contig9253	GL548669	25153	232	1071	-	GL565792	1237	1	1015	TCONS_00187906	-	GL564811	1597	990	1521	+
Contig16407	GL548662	5488	4390	5488	+	GL571417	1381	1	1374	TCONS_00190566	-	GL566046	1107	8	1104	-
Contig2377	GL562747	483419	480322	480493	+	GL570608	1089	1	548	TCONS_00190187	-	GL567455	1247	685	1247	+
Contig42046	GL559925	1872	35	848	+	GL572346	1100	6	1100	TCONS_00191005	-	GL568926	1377	195	349	-
Contig1256	GL549736	5442	4362	4542	+	GL549736	5442	4543	5442	TCONS_00008800	-	GL569202	1632	1	1632	+
Contig9116	GL554036	2307	690	1081	+	GL564676	1061	17	1061	TCONS_00187428	-	GL571023	1111	277	1109	+

Table 1 Contigs from assembly 18 suggesting possible scaffold concatenation

The table reports potentially truncated Cuffmerge contigs (TCONS) located on different scaffolds aligning with a single contig from assembly 18. Names of scaffolds (GL), length, coordinates and orientation of alignment with contig are indicated.

Chapitre 2 : Analyse d'expression des gènes chimiosensoriels entre populations, sexes et tissus, de deux espèces sœurs de *Rhodnius*, *R. prolixus* et *R. robustus*

Résumé : Le genre *Rhodnius* comprend des espèces domiciliées, et d'autres sylvatiques. Parmi elles, deux espèces sœurs, *R. prolixus* et *R. robustus*, très proches morphologiquement et génétiquement, présentent des écotypes différents. Ainsi, *R. prolixus* est strictement domiciliée alors que *R. robustus*, inféodée aux palmiers, visite sporadiquement les habitations en Amazonie. Afin d'apporter quelques éléments d'information sur la capacité d'adaptation de ces deux espèces aux milieux anthropiques, nous avons comparé les transcriptomes chimiosensoriels de souches des deux espèces provenant du même élevage et d'échantillons de *R. robustus* prélevés sur le terrain en Guyane. Le génome de *R. prolixus* a été utilisé comme référence, pour générer un transcriptome commun aux deux espèces. Nous nous sommes focalisés dans un premier temps sur l'étude du répertoire des gènes chimiosensoriels (OBP, CSP, OR, GR, IR) et également des gènes *takeout* puis nous avons effectué une analyse d'expression différentielle. Le nombre de transcrits des gènes chimiosensoriels et des gènes *takeout* exprimés ainsi que les isoformes majoritaires de ces gènes diffèrent en fonction de l'espèce, du sexe et de l'environnement. De même, l'analyse d'expression différentielle révèle des divergences d'expression en fonction du sexe et de l'environnement, notamment des gènes chimiosensoriels, mais aussi de gènes impliqués dans la réponse au stress (protéines heat-shock) et aux agressions extérieures (cytochrome P450, protéines cuticulaires).

Introduction

Les espèces *Rhodnius prolixus* and *R. robustus* présentent une distribution géographique qui se superpose dans les pays du nord de l'Amérique du Sud (Colombie, Venezuela, Bolivie, Equateur, Guyane française et nord du Brésil). Du fait de la difficulté de distinguer ces deux espèces morphologiquement (Harry, 1993a, 1994), la délimitation précise de leurs aires de distribution est rendue difficile. L'espèce *R. prolixus* est considérée comme domiciliée, et bien qu'elle ait été éradiquée au Costa Rica, au Chili, au Guatemala, au Mexique, au Nicaragua, à Salvador, et en Uruguay, (Dias et al., 2002; Hashimoto and Schofield, 2012), elle représente encore aujourd'hui le vecteur principal de la maladie de Chagas en Colombie, au Venezuela et dans certains pays d'Amérique Centrale (WHO, 2015). Cependant, la situation est plus complexe car certains auteurs décrivent des populations sylvatiques de *R. prolixus* au Venezuela (Felicangeli et al., 2007; Fitzpatrick et al., 2008; Sanchez-Martin et al., 2006) et en Colombie (Guhl et al., 2009; Rendón et al., 2015). L'espèce *R. robustus*, quant à elle, est largement distribuée dans la région amazonienne et est considérée comme sylvatique (Abad-Franch and Monteiro, 2007). Cette espèce présente des pourcentages élevés d'infection par *T. cruzi* (Miles et al., 1983). L'invasion sporadique des habitations par ces punaises, apparemment liée à leur attraction par des sources de lumière artificielle, a été décrite plusieurs fois, que ce

soit au Pérou (Cabrera et al., 2013), en Guyane (Bérenger et al., 2009), au Venezuela (Felicangeli et al., 2002), ou au Brésil (Aguilar et al., 2007). Ce phénomène favoriserait la transmission de *T. cruzi* à l'homme (Castro et al., 2010). De plus, en région amazonienne, un autre mode de transmission par voie orale est en constante augmentation (Coura et al., 2002; Monteiro et al., 2010; Valente et al., 1999), impliquant notamment *R. robustus* qui migre des milieux sylvatiques vers les lieux de stockage des fruits de palmiers ou de production de jus à partir de ces fruits.

Ainsi, l'espèce *R. prolixus* mais également *R. robustus* jouent un rôle important dans la transmission de la maladie de Chagas. Cependant, leur écologie divergente nous questionne sur les raisons qui distinguent ces deux espèces sœurs dans leur capacité à coloniser le milieu anthropique. Le couple *R. prolixus*/*R. robustus* constitue donc un bon modèle d'étude du processus de domiciliation au sein du genre *Rhodnius*.

Dans ce contexte, les transcriptomes chimiosensoriels de différentes souches/populations de *R. prolixus* et *R. robustus* ont été comparés. Du fait qu'il n'existe pas de population domiciliée de *R. robustus*, nous avons utilisé une souche d'élevage, pour « mimer » l'adaptation au milieu anthropique. Une souche de *R. prolixus* provenant du même élevage a été utilisée pour faire les comparaisons d'expression génique. Des échantillons de *R. robustus* sauvages prélevés sur le terrain en Guyane ont également été utilisés. Notre objectif était de comparer l'expression des gènes exprimés dans les organes chimiosensoriels, en fonction i) de l'environnement (sauvage/élevage), ii) de l'espèce (*R. robustus*/*R. prolixus*), et iii) du sexe des individus.

Pour pouvoir comparer les deux espèces, un transcriptome de référence commun aux deux espèces a été construit en utilisant une méthode d'assemblage basée sur le génome de *R. prolixus*. Les analyses d'expression différentielle permettent de comparer l'expression d'un gène entre organismes qui diffèrent pour certaines conditions. Cependant, elles ne permettent pas d'analyser les gènes exprimés dans une condition et non exprimés dans une autre. Or, ces divergences peuvent refléter des différences d'adaptation au niveau des organismes. C'est pourquoi, dans un premier temps, une étude comparative de présence/absence de transcrits de gènes chimiosensoriels et de gènes *takeout* a été réalisée. Par la suite, une analyse d'expression différentielle basée sur les gènes communs à *R. robustus* et *R. prolixus* a été effectuée via les packages DESeq2 et EdgeR.

Matériel et Méthodes

Echantillonnage, extraction d'ARN et séquençage

Pour l'espèce *R. prolixus* nous avons utilisé deux souches d'élevage (Tableau 1) : l'une d'un élevage brésilien de l'Insetário de Triatominae da Faculdade de Ciências Farmacéuticas, Universidade Estadual Paulista (UNESP) - Araraquara, (PMEF et PFEF) et l'autre d'un

élevage suisse de l'Institut de Biologie, Université de Neuchâtel (PMES). Pour l'espèce *R. robustus*, nous avons utilisé des punaises récupérées par piège lumineux ou lors de visites sporadiques dans les habitations en Guyane (RFG1, RFG2 et RMG). Nous avons inclus dans notre étude des échantillons de *R. robustus* d'élevage (RFEF et RMEF) provenant du même laboratoire brésilien que les *R. prolixus*, afin de comparer sur l'expression des gènes chimiosensoriels : d'une part l'effet de l'environnement (sauvage/élevage), d'autre part l'effet interspécifique (*R.robustus/R.prolixus*) en utilisant des souches des deux espèces strictement élevées dans les mêmes conditions.

Afin de cibler les gènes du système chimiosensoriel, l'ARN a été extrait à partir du rostre et des antennes des punaises. La tête a été ajoutée à ce protocole pour les échantillons PMES et RFG1 afin de compléter le panel des gènes dans les transcriptomes de référence. L'extraction a été réalisée par TRIzol®. Afin d'estimer l'expression moyenne des gènes au sein de chaque population, chaque échantillon (*R. prolixus* et *R. robustus*) comprend plusieurs individus regroupés par origine et par sexe (à l'exception de l'échantillon RFG1 qui ne comprend qu'un individu). Malheureusement, du fait de difficultés d'échantillonnage, les échantillons de la population de Guyane de *R. robustus* (population sauvage) sont de faible effectif en comparaison des autres. Les échantillons ont ensuite été séquencés en Illumina paired-end par la Plateforme génomique IBENS, Institut de Biologie de l'École Normale Supérieure, Paris, France en suivant le protocole décrit dans Marchant et al., (2015a).

Echantillon	Espèce	Sexes	Environnement	Organes	Nombre d'individus par échantillon	Nombre de paires de reads
PMEF	<i>R. prolixus</i>	M	Elevage Brésil	Antennes + Rostres	18	31 125 782
PFEF	<i>R. prolixus</i>	F	Elevage Brésil	Antennes + Rostres	18	20 372 071
PMES	<i>R. prolixus</i>	M	Elevage Suisse	Tête	5	42 792 624
RFEF	<i>R. robustus</i>	F	Elevage Brésil	Antennes + Rostres	21	30 786 251
RMEF	<i>R. robustus</i>	M	Elevage Brésil	Antennes + Rostres	19	26 748 452
RFG1	<i>R. robustus</i>	F	Sauvage Guyane	Tête	1	34 455 341
RFG2	<i>R. robustus</i>	F	Sauvage Guyane	Antennes + Rostres	2	24 417 527
RMG	<i>R. robustus</i>	M	Sauvage Guyane	Antennes + Rostres	4	32 908 126

Tableau 1: Echantillons de *Rhodnius* et données de séquençage

La première lettre de chaque échantillon indique l'espèce (P pour *prolixus* et R pour *robustus*); la seconde lettre indique le sexe (M pour mâle et F pour femelle) et la troisième lettre indique l'origine des insectes (EF pour les souches en provenance d'élevage brésilien, ES pour les souches en provenance de Suisse et G pour les échantillons sauvages de Guyane)

Réalisation d'un transcriptome de référence commun à *R. prolixus* et *R. robustus*

Contrairement à *R. prolixus*, nous ne disposons pas de données transcriptomiques 454 pour les échantillons de *R. robustus*. Deux types d'assemblage ont été réalisés. Dans un premier temps, les lectures des différents échantillons de *R. robustus* ont été nettoyées et assemblées *de novo* par le biais du logiciel Trinity (pour les paramètres de nettoyage et d'assemblage, consulter Marchant et al., 2015b, chapitre précédent). Dans un second temps des assemblages par alignement sur le génome de *R. prolixus* ont été réalisés en utilisant les données d'annotation du génome et le logiciel d'alignement des lectures Tophat (Trapnell et al., 2009) et l'assembleur Cufflinks (Trapnell et al., 2010) (pour les paramètres d'alignement et d'assemblage, consulter Marchant et al., 2015b).

Le script Cuffmerge du package Cufflinks a ensuite été utilisé pour construire un transcriptome de référence *R. prolixus/R.robustus* regroupant l'ensemble des transcrits de tous les échantillons des deux espèces nommé TRef-RP-A (pour Transcriptome de Référence *R.prolixus/R.robustus* pour l'analyse d'Annotation) (Figure 1). Ce transcriptome TRef-RP-A a été utilisé pour l'annotation, l'analyse de présence-absence, et l'analyse des isoformes. Ainsi, les gènes chimiosensoriels (OBP, CSP, IR, GR, OR) annotés dans le génome de *R. prolixus* ont été recherchés dans ce transcriptome TRef-RP-A. D'autres transcrits (*takeout*, protéines cuticulaires, protéines heat-shock) ont également été annotés par Blast. La présence de lectures alignées en « Reads Per Kilobase per Million mapped reads » ou RPKM (normalisation des données de comptage par la longueur totale des transcrits et le nombre total de lectures alignées), évaluée par le script Cuffdiff du package Cufflinks, a été utilisée pour i) faire un comparatif présence/absence des gènes chimiosensoriels et *takeout* entre nos échantillons et ii) évaluer l'isoforme la plus exprimée (isoforme majoritaire), parmi les isoformes d'un même gène. Pour l'analyse d'expression différentielle, un autre transcriptome de référence a été construit (Figure 1) et nommé TRef-RP-DE (pour Transcriptome de Référence *R.prolixus/R.robustus* pour l'analyse d'Expression Différentielle). Seuls les transcrits présents (détection de lectures alignées) chez les deux espèces de *Rhodnius* ont été gardés pour construire ce transcriptome et une seule isoforme par gène a été conservée en se basant sur la représentativité des différents isoformes (nombre de lectures alignées en RPKM évalué par Cuffdiff).

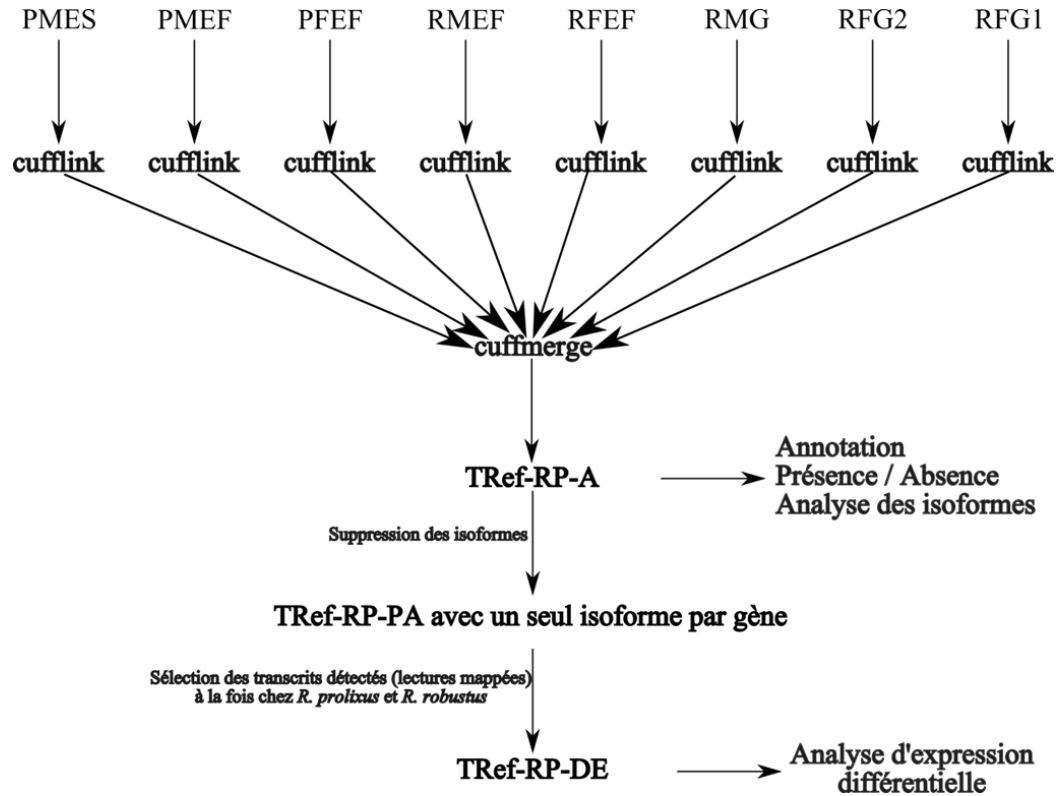


Figure 1 : Workflow représentant les différentes étapes d'assemblage des transcriptomes utilisés pour les différentes analyses. TRef-RP-A est le transcriptome de référence de *R. prolixus* et *R. robustus* utilisé pour l'annotation, l'analyse de la présence / absence et l'analyse de l'utilisation des isoformes des gènes d'intérêt. TRef-RP-DE est le transcriptome de référence de *R. prolixus* et *R. robustus* utilisé pour l'analyse d'expression différentielle)

Les échantillons séquencés à partir de l'ARN des têtes présentent vraisemblablement des profils d'expression différents de ceux obtenus uniquement à partir des antennes et des rostrs. PMES et RFG1 ont donc été enlevés du jeu de données pour réaliser l'analyse d'expression différentielle. Cette dernière a été réalisée à partir de deux packages R : EdgeR (McCarthy et al., 2012) et DESeq2 (Love et al., 2014). En ce qui concerne EdgeR, nous avons choisi la normalisation TMM. La normalisation de DESeq2 est faite automatiquement. Pour EdgeR et DESeq2, les données de comptage sont modélisées par un modèle linéaire généralisé (GLM) basé sur une loi de distribution binomiale négative. Trois facteurs sont pris en compte dans cette analyse : l'espèce, le sexe et l'environnement. L'évaluation de la dispersion par EdgeR a été réalisée en utilisant l'option « robust ». La comparaison des différentes conditions des trois facteurs a été réalisée avec les deux logiciels (*R. prolixus* versus *R. robustus* ; mâle versus femelle ; sauvage versus élevage). La correction de Benjamini-Hochberg (Benjamini and Hochberg, 1995) a été appliquée avec les deux logiciels, les contigs étant considérés comme différentiellement exprimés (DE) lorsque la P-ajustée est inférieure à 0,05. Les contigs définis comme DE ont été annotés par Blast sur la base de données non redondante des protéines (NR) de mai 2015. En parallèle, une comparaison des deux espèces *R. prolixus* versus *R. robustus* d'élevage a été réalisée dans les mêmes conditions.

Résultats

Comparaison des différents assemblages

R. robustus étant une espèce distincte de *R. prolixus*, nous voulions nous assurer qu'un assemblage basé sur la référence du génome de *R. prolixus* serait au moins aussi informatif qu'un assemblage *de novo*. C'est pourquoi nous avons réalisé les deux types d'assemblage. De la même manière que pour les assemblages de *R. prolixus* présentés au chapitre précédent, l'assemblage de *R. robustus* basé sur le génome de référence par Cufflinks augmente considérablement le N50 (compris entre 907 à 1847pb) ainsi que la complétude du transcriptome (96,77% à 98,39% du jeu de protéines CEGMA retrouvé) par rapport à un assemblage *de novo* (N50 compris entre 499 et 749pb et seulement 6,45 à 12,90% du jeu de protéines CEGMA retrouvé ; Tableau 2). C'est pour quoi nous avons sélectionné la méthode d'assemblage basée sur la référence pour constituer notre transcriptome commun.

Les assemblages Cufflinks des différents échantillons de *R. prolixus* et de *R. robustus* ont par la suite été fusionnés grâce au script Cuffmerge. Cet assemblage combiné rassemble 184 271 transcrits utilisés pour l'annotation (TRef-PR-A). Pour chaque locus, seule l'isoforme la plus représentative de l'ensemble de nos échantillons (somme des RPKM) a été sélectionnée, ce qui réduit notre transcriptome à 136 988 transcrits. Sur l'ensemble de ces 136 988 transcrits, 127 723 (soit 91%) sont exprimés dans au moins un échantillon de *R. prolixus* et au moins un échantillon de *R. robustus* et ont été retenus pour constituer le transcriptome TRef-PR-DE utilisé pour l'analyse différentielle.

Assemblages	Nombre de contigs	Longueur totale	Distribution de taille (bp)			CEGMA (%)	
			max	moyenne	N50 (bp)	complet	partiel
RFEF Trinity	197 795	98 047 341	5 681	495,70	582	1.61	7.26
RFG1 Trinity	261 769	155 727 499	22 859	594.90	749	NA	NA
RFG2 Trinity	174 031	79 282 047	11 764	455,56	499	5.24	12.90
RMEF Trinity	619 061	297 368 856	7 742	480,35	552	NA	NA
RMG1 Trinity	144 371	69 083 365	5 386	478,51	546	0.81	6.45
RFEF Cufflink	131 760	125 817 722	248 692	954.90	1 246	86.29	97.18
RFG1 Cufflink	93 238	124 846 519	248 692	1339.01	1 847	91.53	98.39
RFG2 Cufflink	105 493	69 951 034	248 692	663.09	907	85.48	97.18
RMEF Cufflink	205 727	328 748 578	248 692	1597.98	2 091	88.71	97.58
RMG1 Cufflink	45 620	36 842 101	248 692	807.59	1 451	84.68	96.77
TRef-PR-A Cuffmerge	184 271	483 000 356	256 221	2621.14	4 318	93.55	98.79
TRef-PR-A Cuffmerge Une isoforme/locus	136 988	338 277 756	256 221	2469.40	3 885	71.77	89.92
TRef-PR-DE Cuffmerge	127 723	331 831 140	256 221	2598.05	3 920	67.74	85.89

Tableau 2 : Récapitulatif des assemblages *de novo* (Trinity), des assemblages basés sur la référence du génome de *R. prolixus* (Cufflink) des différents échantillons de *R. robustus*, et des assemblages des transcriptomes de référence *R. prolixus*/*R. robustus* (Cuffmerge), TREF-PR-A, utilisé pour l'annotation et TRefPR-DE, utilisé pour l'analyse d'expression différentielle. Le nombre de contigs, le nombre total de bases, la taille du plus grand contig (max), la taille moyenne des contigs, le N50 (longueur des contigs à partir de laquelle la somme des longueurs des contigs de même taille ou plus grands couvre 50 % de l'assemblage) et les résultats du CEGMA (le logiciel « Core eukaryotic genes dataset » évalue la complétude d'un génome ou d'un transcriptome à partir d'un set de 248 protéines ultra conservées chez l'ensemble des eucaryotes. « Complet » désigne le pourcentage de protéines de ce set dont la séquence a été retrouvée complète. « Partiel » représente le pourcentage de protéines de ce set dont au moins 70% de la longueur de la séquence a été identifiée dans le transcriptome).

Présence/absence des transcrits des protéines de liaisons des molécules odorantes et des transcrits *takeout*

Afin de déterminer si les différents gènes annotés comme chimiosensoriels (CSPs, OBPs, ORs, IRs et GRs) ou *takeout* (To) sont exprimés dans nos échantillons, nous avons regardé la présence de lectures alignées sur ces gènes à partir des données de comptage (RPKM) de Cufflink (Annexe 2, Tableau 3). Nous parlerons dans ce paragraphe de transcrits « détectés » lorsque la présence de lectures alignées est avérée (RPKM > 0). Dans le transcriptome TRef-PR-A, l'ensemble des transcrits OBPs et des CSPs correspondant aux gènes répertoriés dans le génome de *R. prolixus* ont été retrouvés. Sur les 116 ORs, 33IRs et 28GRs répertoriés dans le génome de *Rhodnius prolixus*, 77ORs, 27IRs, 26GRs ont été annotés dans le transcriptome

TRef-PR-A. Dans ce même transcriptome, nous avons également annoté 24 transcrits codant des protéines *takeout*. Au total 200 transcrits ont donc été annotés et étudiés (Tableau 3).

Afin de comprendre dans quelles conditions (espèce, tissus, sexe, environnement) les transcrits de ces familles sont détectés, différentes comparaisons ont été effectuées.

Tout d'abord, nous avons évalué le nombre de transcrits détectés au sein des espèces *R. prolixus* et *R. robustus* quels que soient les organes utilisés (Tableau 3, colonnes 1 et 2). Sur 200 gènes d'intérêt annotés, 194 sont partagés par les deux espèces, soit 97% de transcrits chimiosensoriels et *takeout* communs. Concernant les CSPs, IRs et GRs et To nous avons détecté aussi bien chez *R. prolixus* que *R. robustus* l'ensemble des transcrits répertoriés dans le transcriptome global TRef_PR_A (Tableau 3). Concernant les OBPs, sur les 28 annotés dans le génome de *R. prolixus*, le transcrit OBP2 n'a été retrouvé dans aucune des deux espèces tandis que l'OBP9 n'est détecté que chez *R. prolixus*. Pour les ORs sur les 76 gènes recherchés, 73 et 75 transcrits sont retrouvés chez *R. prolixus* et *R. robustus* respectivement. Afin de s'affranchir du facteur environnemental sur la détection des transcrits, nous pouvons comparer *R. prolixus* et *R. robustus* au sein d'une seule et même condition, à savoir les échantillons d'élevage extraits à partir des organes chimiosensoriels (Tableau 3, colonne 3 et 12). Dans ces conditions, la CSP19 détectée chez *R. robustus* est absente chez *R. prolixus*. Même si un nombre similaire d'ORs est retrouvé chez *R. prolixus* (72) et chez *R. robustus* (71), elles diffèrent de par l'expression de 7 ORs. De plus, le transcrit GR13 n'est pas exprimé chez *R. robustus*.

Afin d'évaluer le nombre de transcrits d'intérêt détectés dans les différents organes extraits, nous avons comparé les échantillons extraits à partir des organes chimiosensoriels seuls (antennes et rostre - Tableau 3, colonnes 5) et ceux extraits à partir de la tête entière (tête, antennes et rostre - Tableau 5, colonnes 8). Certains transcrits chimiosensoriels ne sont pas détectés dans les échantillons issus de la tête entière alors qu'on les retrouve dans ceux issus des organes chimiosensoriels seuls. C'est le cas des transcrits des CSPs 14, 15 et 18, des OBPs 9 et 15, de IR41c, et To22. Il est à noter que 8 transcrits d'OR ne sont pas non plus détectés dans les échantillons « tête ». Si ces différents gènes sont faiblement ou pas exprimés dans le cerveau, il est probable que l'ajout des ARN du cerveau à ceux des antennes et des rostre ait abouti à la « dilution » de ces transcrits, ce qui expliquerait l'absence de leur détection dans les échantillons PMES et RFG1. Ce phénomène de dilution est d'autant plus probable pour des gènes déjà très faiblement exprimés dans les antennes et le rostre (CSP14 et 15 par exemple chez *R. prolixus* - Annexe Tableau 2).

Nous avons ensuite étudié la détection des transcrits d'intérêt en fonction du sexe dans les organes chimiosensoriels, quelle que soit l'espèce (Tableau 3, colonnes 9 et 10). Concernant les OBPs, trois transcrits ne sont pas détectés chez les femelles et deux chez les mâles. De même, deux transcrits d'OR ne sont pas détectés chez les femelles et deux autres chez les mâles. Enfin pour les IR, GR et To, un transcrit n'est pas retrouvé dans l'un des sexes.

Enfin, nous avons comparé les transcrits détectés selon les conditions de vie des insectes quels que soient le sexe et l'espèce, en analysant les insectes d'élevage et sauvages (Tableau 3, colonne 11 et 13). Plusieurs transcrits sont détectés spécifiquement dans les échantillons d'élevage. C'est le cas de transcrits de trois CSPs, cinq OBPs, six ORs, un IR, un GR ainsi que deux To. A l'inverse, seul l'OR22, présent dans les échantillons sauvages n'est pas détecté dans les échantillons d'élevage. Cependant, l'environnement d'élevage rassemble les échantillons des deux espèces alors que seule *R. robustus* est représentée dans l'environnement sylvatique. Cette comparaison pourrait donc être biaisée car révélant certains transcrits propres à *R. prolixus*. La comparaison des échantillons de *R. robustus* seul (en condition d'élevage colonne 12 et en condition sylvatique colonne 13) montre d'autres transcrits spécifiques aux *R. robustus* sauvages : c'est le cas de 3 ORs, et 1 GR.

Sur l'ensemble des 200 transcrits chimiosensoriels et *takeout* annotés, 5 divergent entre les espèces, 15 entre les organes chimiosensoriels par rapport à la tête, 10 entre sexes et 19 selon l'environnement (élevage ou sauvage).

Cependant, parmi ces transcrits, beaucoup ont peu de lectures alignées dans l'échantillon où ils sont détectés (RPKM<1). L'absence de détection dans un échantillon pourrait être liée au seuil de détection. Notons cependant des différences notables du nombre de lectures pour les transcrits OBP18, OR20, GR 24 (retrouvés uniquement chez les mâles) IR75b (détecté uniquement chez les femelles), ainsi que pour les transcrits OBP19, OBP22, OR12, OR43, IR75b, IR75k et GR16, en liaison avec l'environnement (Annexe Tableau 2).

Partie II : Analyse des transcriptomes chimiosensoriels chez différentes espèces du genre *Rhodnius*

	1	2	3	4	5	6	7	8	9	10	11	12	13
Gènes (nbr annotés)	<i>R. prolixus</i> (PMEF +PFEF +PMES)	<i>R. robustus</i> (RFEF +RMEF +RFG1 +RFG2 +RMEF)	Organes chimioS <i>R. prolixus</i> (PMEF +PFEF)	Organes chimioS <i>R. robustus</i> (RFEF +RMEF +RFG2 +RMG)	Total Organes chimioS (PMEF +PFEF +RFEF +RMEF +RFG2 +RMG)	Tête <i>R. prolixus</i> (PMES)	Tête <i>R. robustus</i> (RFG1)	Total Tête (PMES +RFG1)	Organes chimioS Femelles (PFEF +RFEF +RFG2)	Organes chimioS Mâles (PMEF +RMEF +RMG)	Total Organes chimioS Elevages (PMEF +PFEF +RFEF +RMEF)	Organes chimioS Elevages <i>R. robustus</i> (RFEF +RMEF)	Organes chimioS <i>R. robustus</i> Sauvages (RFG2 +RMG)
CSP (19)	19	19	18 (CSP19)	19	19	16 (CSP14, 15, 18)	16 (CSP14, 15, 18)	16 (CSP14, 15, 18)	19	19	19	19	16 (CSP15, 17, 19)
OBP (28)	27 (OBP2)	26 (OBP2, 9)	27 (OBP2)	26 (OBP2, 9)	27 (OBP2)	25 (OBP2, 9, 15)	25 (OBP2, 9, 15)	25 (OBP2, 9, 15)	25 (OBP2, 7, 18)	26 (OBP2, 9)	27 (OBP2)	26 (OBP2, 9)	22 (OBP2, 7, 9, 15, 19, 22)
OR (76)	73 (OR21, 22, 60)	75 (OR89)	72 (OR21, 22, 60, 75)	74 (OR72, 89)	76	52 (OR3, 5, 10, 12, 16, 18, 21, 22, 23, 30, 42, 47, 49, 53, 54, 57, 60, 61, 70, 86, 89, 91, 97, 104)	58 (OR5, 18, 21, 22, 31, 49, 52, 60, 64, 68, 71, 80, 81, 82, 88, 89, 91, 114)	68 (OR5, 18, 21, 22, 49, 60, 89, 91)	74 (OR20, 21)	74 (OR72, 89)	75 (OR22)	71 (OR22, 23, 72, 89, 90)	70 (OR12, 21, 43, 64, 72, 89)
IR (27)	27	27	26 (IR75b)	27	27	25 (IR41c, 75n)	26 (IR41c)	26 (IR41c)	27	26 (IR75b)	27	27	26 (IR75k)
GR (26)	26	26	26	26	26	21 (GR15, 16, 19, 21, 25)	25 (GR 21)	26	25 (GR24)	26	26	25 (GR13)	25 (GR16)
To (24)	24	24	24	24	24	22 (To2, 22)	22 (To1, 22)	23 (To22)	23 (To16)	24	24	24	22 (To16, 22)
200	196	197	193	196	199	161	171	184	193	195	198	192	181

Tableau 3 : Nombre de transcrits chimiosensoriels et *takeout* détectés (présence de lectures cartographiées) dans les différents échantillons regroupés par espèce, tissu, sexe ou environnement. Le nombre de transcrits est indiqué dans chaque case. Les noms des transcrits non détectés sont indiqués entre parenthèses.

Isoformes majoritaires des protéines de liaisons des molécules odorantes et des gènes *takeout*

Lorsque l'assembleur Cuffmerge a défini plusieurs isoformes à partir d'un même gène, nous avons considéré comme isoforme majoritaire celle comportant le plus de lectures alignées. En Annexe Tableau 3, un code couleur montre pour les différentes comparaisons (espèce, sexe, environnement) si l'isoforme trouvée diffère ou non de cette isoforme majoritaire.

Les résultats résumés par famille de gènes et par comparaison entre les conditions sont donnés dans le Tableau 4. L'isoforme majoritaire peut varier selon l'espèce, les tissus, le sexe ou l'environnement (Annexe, Tableau 3). En fonction de l'espèce, l'isoforme majoritaire diffère pour 66 transcrits sur 188 (32%) dans les organes chimiosensoriels (antennes et rostre), et pour 45 transcrits sur 129 (35%) dans la tête. En fonction du sexe, l'isoforme majoritaire diffère pour 57 transcrits sur 184 (31%). Enfin, entre environnements, l'isoforme majoritaire diffère pour 73 transcrits sur 177 (41%). Il est à noter que les transcrits *takeout* présentent le plus de variation d'isoformes. Les isoformes majoritaires des CSPs et des GRs sont les moins variables. Les gènes des IRs d'une part et des ORs d'autre part (OrCo inclus) présentent des isoformes majoritaires

particulièrement différentes entre les échantillons des deux espèces extraits respectivement des organes chimiosensoriels et des têtes.

	ChimioS Rpro/Rrob			Tête Rpro/Rrob			M/F			Elev/Sauv		
	nb	%	Total	nb	%	Total	nb	%	Total	nb	%	Total
CSP	2	11,11	18	2	13,33	15	3	15,79	19	2	12,5	16
OBP	5	20	25	6	27,27	22	7	31,82	22	9	42,9	21
OR	24	34,29	70	16	53,33	30	23	32,86	70	33	47,8	69
OrCo	0	0	0	1	100	1	0	0,00	0	0	0	0
IR	14	56	25	9	39,13	23	9	36,00	25	10	41,7	24
GR	4	15,38	26	2	10,53	19	4	16,00	25	6	24	25
To	12	50	24	9	47,37	19	11	47,83	23	13	59,1	22
Total	61	32,45	188	45	34,88	129	57	30,98	184	73	41,2	177

Tableau 4 : Nombre de gènes par famille multigénique présentant une isoforme majoritaire différente entre échantillons au sein de la condition analysée avec nb : nombre, % pourcentage tenant compte du nombre total de gènes au sein de la famille considérée (seuls les gènes où il y avait un transcrit pour chaque condition ont été pris en compte).

Recherche de gènes différentiellement exprimés

Nous avons réalisé une analyse d'expression différentielle prenant en compte trois facteurs : l'espèce (*R. prolixus* / *R. robustus*), le sexe et l'environnement (élevage ou sauvage).

Pour la construction du transcriptome utilisé comme référence pour les analyses d'expression différentielle (TRef_PR_DE) nous n'avons utilisé qu'une isoforme par gène (celle détectée comme majoritaire sur l'ensemble des échantillons). Par ailleurs, si une isoforme n'était pas détectée dans une des espèces, elle a été exclue de cette analyse. Ainsi, sur le total des gènes chimiosensoriels annotés, 180 ont été analysés (18 CSPs, 24 OBP, 67 ORs, 25 IRs, 24 GRs, 22 Tos).

La Figure 2 montre l'ACP réalisée à partir des données de comptage normalisées grâce à la fonction rlog de DESeq2. Deux facteurs expliquent 84% de la variabilité observée : l'environnement et le sexe. A première vue, il semblerait que la différence entre sexes soit accentuée dans l'environnement sauvage. Le regroupement des échantillons en fonction de l'expression des 30 contigs les plus différentiellement exprimés entre sauvage et élevage (heatmap) confirme les observations données par l'ACP, notamment en ce qui concerne l'échantillon RMG1 qui s'écarte d'avantage des autres échantillons (Figure 2).

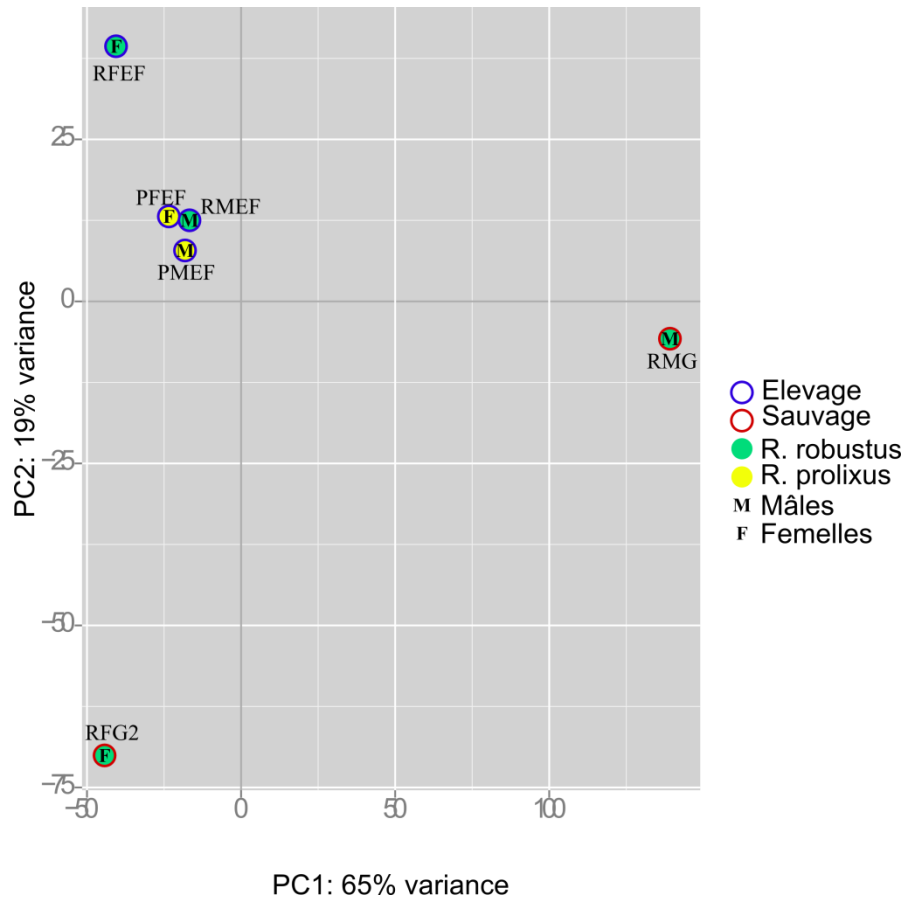


Figure 2 : ACP basée sur les données de comptage normalisées par la fonction rlog de DESeq2.

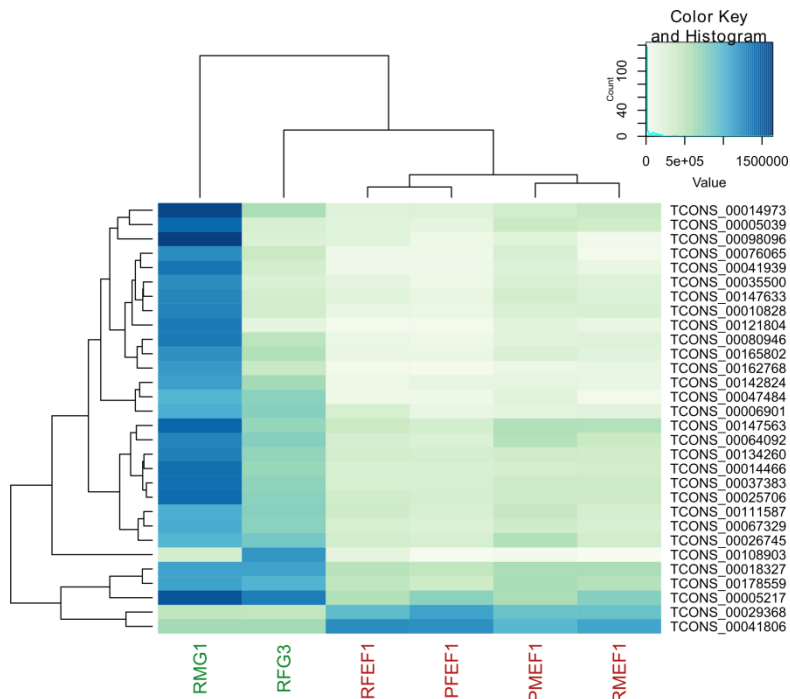


Figure 3 : Heatmap réalisée avec DESeq2 indiquant les regroupements des différents échantillons en fonction des données d'expression des 30 contigs les plus différentiellement exprimés (en fonction de la P-ajustée) entre sauvage (vert) et élevage (rouge).

Le graphique de dispersion présente une allure typique de dispersion entre réplicats biologiques en données RNAseq : la dispersion diminue lorsque l'expression augmente. Aucune valeur aberrante n'est visible (Figure 3).

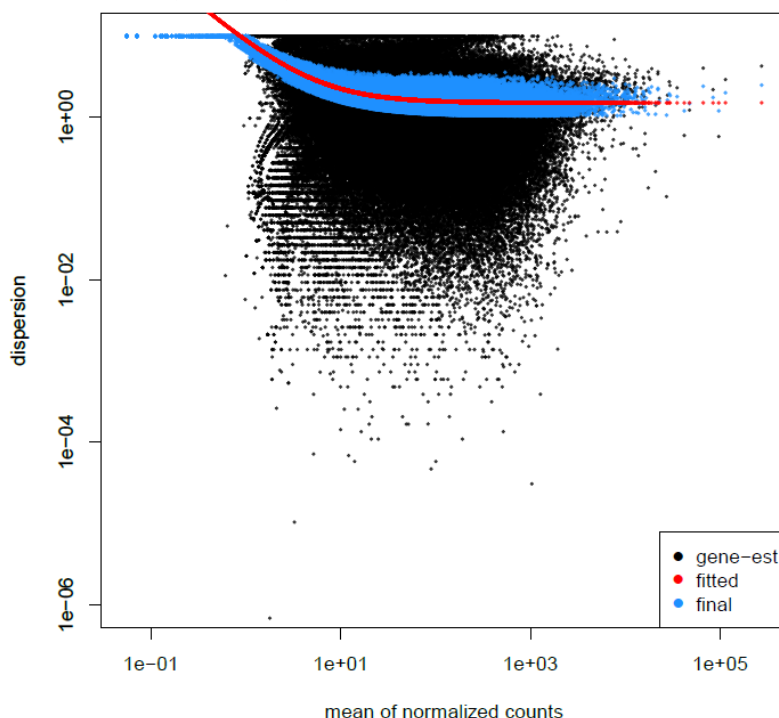


Figure 3 : Dispersion des mesures de comptage de gènes en fonction de l'expression moyenne évaluée par DESeq2. On distingue sur ce graphe la dispersion de chaque transcrit (noir), la courbe de tendance pour tous les échantillons (rouge) permettant d'estimer la valeur corrigée de la dispersion pour chaque transcrit (bleu).

Malgré le fait que DESeq2 et EdgeR soient tous deux basés sur la loi binomiale négative et que la majorité des transcrits détectés DE par EdgeR le soient également par DESeq2, les deux packages diffèrent drastiquement sur le nombre de gènes différentiellement exprimés détectés. Ainsi, DESeq2 évalue 7 à 260 fois plus de transcrits DE que EdgeR suivant la comparaison réalisée (Tableau 4). Cet écart important pourrait être imputé aux méthodes d'évaluation de la dispersion ou à celles de la normalisation qui diffèrent entre les deux packages.

Notons qu'aucun transcrit DE n'a été détecté entre les deux espèces, ni par le logiciel DESeq2 ni par EdgeR. Le même constat peut être fait suite aux analyses d'expression différentielles entre les deux espèces où seuls les *R. robustus* d'élevage ont été pris en compte. En revanche, 145 (EdgeR) à 1012 (DESeq2) transcrits ont été trouvés comme DE entre les sexes avec une tendance globale pour les mâles à sur-exprimer ces transcrits. DESeq2 a détecté 6253 transcrits DE entre les échantillons d'élevage et sauvages alors que EdgeR n'en compte que 24.

	DESeq2			EdgeR			Commun à DESeq2 et EdgeR		
	+	-	Tot	+	-	tot	+	-	tot
Rpro_élevage/Rrob_élevage	0	0	0	0	0	0	0	0	0
Rpro/Rrob	0	0	0	0	0	0	0	0	0
M/F	935	77	1012	143	2	145	140	2	142
Elev/Sauv	2982	3271	6253	3	11	24	1	8	9

Tableau 4 : Tableau récapitulatif des analyses d'expression différentielle évaluées par DESeq2 et EdgeR sur trois composantes : le sexe (M/F pour mâles *versus* femelles), l'espèce (Rpro/Rrob pour *R. prolixus* *versus* *R. robustus*), et l'environnement (Elev/Sauv pour élevage *versus* sauvage).

Parmi les contigs retrouvés DE avec les deux logiciels, très peu ont pu être annotés par Blast (seuls 31 contigs parmi les 142 DE de la comparaison « mâles/femelles »). Par ailleurs, ces quelques contigs annotés sont principalement des transcrits impliqués dans le métabolisme général (protéine gag, polymérase, endonucléase...), des rétrotransposons ou des protéines prédites dont la fonction reste indéterminée.

Par conséquent, nous avons également choisi d'annoter les contigs évalués comme DE par DESeq2 uniquement (Tableau 5).

Ainsi sur 180 gènes chimiosensoriels, seul le GR21 a pu être identifié comme sous-exprimé chez les femelles par rapport aux mâles et onze sont considérés comme DE entre les échantillons élevés en laboratoire et ceux récoltés en milieu sauvage : trois CSPs et une OBP (toutes sous-exprimées dans les échantillons d'élevage), six ORs (dont quatre sur-exprimés et deux sous-exprimés dans les échantillons d'élevage), deux IRs (tout deux sous-exprimés dans les échantillons d'élevage) et un GR (sur-exprimé dans les échantillons d'élevage). En revanche, aucun transcrit de *takeout* n'a été retrouvé DE dans cette analyse. Nous nous sommes également focalisés sur d'autres transcrits qui diffèrent entre les deux conditions environnementales et qui sont impliqués dans des mécanismes de détoxification ou de résistance à la dessiccation. Ainsi, l'annotation des gènes DE révèle pour la comparaison mâles/femelles un transcrit de protéines cuticulaires, sous-exprimé chez les femelles et pour la comparaison échantillons élevage/sauvages les transcrits de : trois cytochromes P450, trois heat-shock protéines et sept protéines cuticulaires.

Contig	Annotation	Comparaison DESeq	Log2 fold	Padj
TCONS_00061919	CSP1	Elev/Sauv	-4.426	0.042
TCONS_00061922	CSP2	Elev/Sauv	-5.222	0.032
TCONS_00069224	CSP11	Elev/Sauv	-6.791	0.006
TCONS_00177364	OBP23	Elev/Sauv	-5.977	0.007
TCONS_00034389	Or40	Elev/Sauv	4.912	0.018
TCONS_00056495	Or13	Elev/Sauv	5.411	0.025
TCONS_00059353	Or33	Elev/Sauv	-5.671	0.004
TCONS_00086863	Or57	Elev/Sauv	-8.584	0.001
TCONS_00143424	Or14	Elev/Sauv	-5.053	0.025
TCONS_00164670	Or10	Elev/Sauv	-5.882	0.027
TCONS_00102111	Ir41c	Elev/Sauv	-5.136	0.018
TCONS_00151556	Ir75e	Elev/Sauv	-4.305	0.046
TCONS_00058943	Gr26	Elev/Sauv	6.65	0.031
TCONS_00158462	Gr21	F/M	-5.855	0.046
TCONS_00126583	cytochrome P450 CYP44a	Elev/Sauv	-4.617	0.042
TCONS_00009216	cytochrome P450 CYP4GGG	Elev/Sauv	5.206	0.040
TCONS_00053742	cytochrome P450 CYP6f	Elev/Sauv	6.295	0.045
TCONS_00100030	70 kDa heat shock protein, partial <i>Pyrrhocoris apterus</i>	Elev/Sauv	-7.120	0.008
TCONS_00069056	heat shock protein 40 <i>Locusta migratoria</i>	Elev/Sauv	-6.992	0.002
TCONS_00149534	heat shock protein 90 <i>Apolygus lucorum</i>	Elev/Sauv	-5.451	0.007
TCONS_00120243	cuticular protein PxutCPR107 <i>Papilio xuthus</i>	Elev/Sauv	-5.929	0.031
TCONS_00066718	larval cuticle protein 1-like, partial <i>Diaphorina citri</i>	Elev/Sauv	4.475	0.049
TCONS_00131308	cuticular protein PpolCPR67A <i>Papilio polytes</i>	Elev/Sauv	-6.583	0.009
TCONS_00021984	cuticular protein RR-2 motif 135 precursor <i>Bombyx mori</i>	Elev/Sauv	-4.394	0.025
TCONS_00071934	cuticular protein CPG12-like precursor <i>Acyrtosiphon pisum</i>	Elev/Sauv	-4.434	0.033
TCONS_00125737	cuticular protein RR-2 family member 7 precursor <i>Nasonia vitripennis</i>	Elev/Sauv	-5.716	0.026
TCONS_00011505	cuticular protein PxutCPR107 <i>Papilio xuthus</i>	Elev/Sauv	-7.825	0.0003
TCONS_00131308	cuticular protein PpolCPR67A <i>Papilio polytes</i>	F/M	-6.682	0.012

Tableau 5: Contigs détectés différentiellement exprimés par DESeq2 et annotés comme appartenant aux familles multigéniques des ORs, IRs, OBPs, CSPs, cytochrome P450, protéines cuticulaires et protéines heat shock

Discussion

R. prolixus, *R. robustus*, ressemblances et divergences

R. prolixus et *R. robustus* étant très proches morphologiquement, leur distinction peut être problématique (Harry, 1993a). En effet, elles se distinguent morphologiquement uniquement par la couleur des pattes des larves les plus âgées et par quelques différences au niveau des plaques basales de l'appareil génital masculin (Lent and Wygodzinsky, 1979). La taille et la forme de leur ailes présentent des asymétries significativement différentes entre les deux espèces (Márquez and Saldamando-Benjumea, 2013). L'absence de marqueurs alloenzymatiques diagnostiques, pour des populations vénézuéliennes de *R. prolixus* et *R. robustus*, avait ainsi conduit Harry, (1993b) à suggérer que *R. robustus* du Venezuela n'était pas un taxon valide. Sur la base de gènes mitochondriaux (*cytochrome b*, 18S rDNA), Lyman et al., (1999) montrent que ces deux espèces sont bien des lignées distinctes. Cependant, les études de Monteiro et al., (2000, 2003) qui prennent en compte différentes populations du bassin amazonien, montrent que *R. robustus* serait un complexe d'espèces cryptiques et que la position taxinomique de *R. prolixus* and *R. robustus* dans le bassin vénézuélien de l'Orénoque reste incertaine.

Notre étude a révélé 91% de transcrits sur 136 988 examinés communs à *R. prolixus* et *R. robustus*. Par ailleurs, la majorité des gènes chimiosensoriels (97%) sont détectés comme exprimés dans les deux espèces de *Rhodnius*. Ces résultats confortent cette forte proximité entre les deux espèces. Pour comparaison, dans une étude menée sur deux espèces sœurs d'abeille récemment séparées, *Euglossa dilemma* et *E. viridissima*, dans un transcriptome construit par méta-assemblage regroupant l'ensemble des transcrits des deux espèces, seuls 80% des transcrits chimiosensoriels annotés sont partagés entre les deux espèces (Brand et al., 2015). Gschloessl et al., (2013) ont étudié le transcriptome issu de la tête et du thorax de deux espèces sœurs, *Ostrinia nubilalis* et *O. scapularis*. Ils ont montré que 80% des transcrits de *Ostrinia nubilalis* s'alignaient par blast sur le transcriptome de *O. scapularis* et réciproquement. Parmi ces transcrits similaires, moins de 10% présentent des identités de séquences supérieures à 80%. De plus, notre analyse d'expression différentielle ne révèle aucune différence significative entre *R. prolixus* et *R. robustus*. Pour comparaison, une analyse similaire entre échantillons d'antennes d'*An. gambiae* et *An. quadriannulatus* révèle un nombre de gènes différentiellement exprimés important (soit 28% et 29,6% de gènes surexprimés chez *An. gambiae* et chez *An. quadriannulatus* respectivement ; Hodges et al., 2014). Ainsi, le nombre de transcrits commun aux deux espèces (notamment ceux codant des protéines chimiosensorielles), et l'analyse d'expression différentielle attestent de la forte similitude qui lie les deux espèces sœurs *R. prolixus* et *R. robustus*.

Néanmoins, deux transcrits (OBP9 et OR89) annotés dans le génome de *R. prolixus* semblent être spécifiques à cette espèce puisqu'ils ne sont pas détectés dans les tissus que nous

avons étudiés de *R. robustus*. Cependant, le nombre très faible de lectures alignées sur ces transcrits montre leur faible expression, ce qui pourrait expliquer leur absence de détection chez *R. robustus*. Notons que trois transcrits ORs et une OBP, bien qu'annotés dans le génome de *R. prolixus*, ne sont pas retrouvés dans les tissus étudiés de *R. prolixus*. Il est tout à fait envisageable que ces gènes soient exprimés dans d'autres tissus connus pour la perception chimiosensorielle, tels que les pattes. C'est probablement le cas pour le transcrit OBP2 absent de tous les échantillons.

Par ailleurs, les deux espèces sœurs présentent jusqu'à 35% d'isoformes majoritaires divergentes. Ces différences sont particulièrement importantes au sein des familles multigéniques des IRs (56%) et des *takeout* (50%). Ces analyses suggèrent que ces deux espèces, bien que très proches génétiquement, et exprimant leurs gènes de manière globalement semblable pourraient néanmoins différer par leurs combinaisons d'isoformes et par leur spécificité d'expression entre tissus. Une autre étude comparative d'espèces sœurs a révélé des différences d'usage d'isoformes entre les deux espèces étudiées (Zhang et al., 2014a). Ainsi, deux espèces sœurs de peupliers du désert ont été étudiées : *Populus euphratica* associée à des sources d'eau souterraines profondes et *P. pruinosa* associée à des sources superficielles à concentration forte en salinité. Elles présentent 90% de transcrits commun mais diffèrent par l'épissage alternatif de plusieurs locus en condition de stress salin. Ainsi, on peut supposer que des phénomènes d'épissage alternatif sont impliqués dans l'adaptation d'une espèce à un nouvel environnement ; comme l'adaptation de *P. pruinosa* à un milieu salin ou l'adaptation de *R. prolixus* à un milieu anthropique.

Mâles/Femelles, divergence d'expression et d'isoformes

Un nombre important de gènes ont été déterminés comme différentiellement exprimés entre les deux sexes et sont probablement impliqués dans le dimorphisme sexuel ou associés à des comportements sexuels spécifiques (Tableau 4). Cependant, la majorité des transcrits DE annotés sont impliqués dans le métabolisme global. Seul un récepteur gustatif est retrouvé. Les autres gènes chimiosensoriels ne semblent pas présenter de différence d'expression mâles-femelles. Ceci est d'autant plus étonnant que de nombreuses études révèlent d'importantes différences d'expression de gènes chimiosensoriels; que ce soit chez les fourmis *C. floridanus* et *H. saltator* (Zhou et al., 2012), le moustique *Anopheles gambiae* (Pitts et al., 2011), ou encore le papillon *Sesamia nonagrioides* (Acín et al., 2009). L'absence de gènes chimiosensoriels différentiellement exprimés entre mâles et femelles pour les 2 espèces de *Rhodnius* étudiées pourrait provenir d'une variabilité d'expression pour ces gènes trop importante entre nos échantillons, ou d'une expression trop faible pour être comparée par nos méthodes.

Cependant, parmi les gènes codant des protéines du système chimiosensoriel ou des protéines de la famille *takeout*, plusieurs semblent être exprimés de manière sexe spécifique. Ainsi, deux OBPs, quatre ORs, une IR et une To ont été détectés uniquement chez les mâles ou chez les

femelles. D'autres espèces présentent également des gènes chimiosensoriels exprimés spécifiquement dans un seul sexe. C'est le cas par exemple des GRs 32a et 68a exprimés uniquement chez les mâles de la drosophile et nécessaires à un comportement de séduction efficace (Miyamoto and Amrein, 2008 ; Bray and Amrein, 2003). L'OBP99b de la drosophile (appelée également *tsx*) est exprimée au niveau du tissu adipeux des mâles mais pas des femelles; tandis que le gène OBP99a présente les caractéristiques inverses (Fujii, 2002). Ces deux gènes sexe spécifiques seraient impliqués dans la perception des phéromones et la modulation du comportement de reproduction chez la drosophile. Chez cette même espèce, plusieurs protéines *takeout* ont été identifiées comme exprimées au niveau de la tête exclusivement chez les mâles (Dauwalder et al., 2002). Chez *Bombyx mori*, le récepteur olfactif OR-1 est exclusivement exprimé par les mâles et est impliqué dans la perception de phéromone sexuelle (Sakurai et al., 2004). Par ailleurs, notre analyse révèle également que 31% des gènes chimiosensoriels et *takeout* présentent des isoformes majoritaires différents entre les deux sexes. Ainsi, malgré l'absence de gènes chimiosensoriels ou *takeout* différenciellement exprimés entre les deux sexes chez *Rhodnius*, certains de ces gènes sont exprimés spécifiquement dans un seul sexe et d'autres présentent une divergence d'isoformes. Ces différences pourraient être révélatrices de gènes ou isoformes impliqués dans les comportements de parade et d'accouplement chez *Rhodnius*.

L'environnement, facteur déterminant

Les individus de *R. robustus* prélevés sur le terrain montrent une expression divergente de ceux de *Rhodnius* d'élevage, qu'ils appartiennent à la même espèce ou à l'espèce *R. prolixus*. Ces différences d'expression concernent aussi bien i) la présence/absence de transcrits, ainsi 15 transcrits chimiosensoriels et deux *takeout* ne sont pas détectés dans les antennes et les rostrs des échantillons sauvages, ii) le type d'isoforme majoritaire et ce pour 41% des transcrits d'intérêt, iii) l'expression différentielle de certains gènes, 13 transcrits DE appartiennent au système chimiosensoriel dont 10 sous-exprimés en conditions d'élevage. Ces différences d'expression des gènes chimiosensoriels pourraient être liées principalement à la recherche alimentaire et/ou de partenaires sexuels comme cela a été démontré pour d'autres modèles (Fox et al., 2001 ; Sánchez-Gracia et al., 2009). En effet, en élevage, contrairement au milieu sylvatique, les punaises sont nourries régulièrement et les partenaires sexuels sont à proximité. De plus, les individus prélevés en milieu sylvatique ont été collectés par des pièges lumineux ou lors de visites sporadiques dans les habitations humaines. Ce comportement démontre une recherche active (de nourriture ou de partenaire) de ces individus en lien avec la surexpression de leurs gènes chimiosensoriels. Par ailleurs, ces différences pourraient également être imputées aux environnements radicalement différents qui distinguent les individus sauvages des individus d'élevages. En effet, des modifications d'expression des gènes chimiosensoriels entre des individus se développant dans des milieux sauvages et d'autres évoluant dans des milieux

anthropisés ont déjà été mis en évidence. C'est le cas par exemple de *Sesamia nonagrioides*, chez qui l'analyse d'expression différentielle entre une population évoluant sur des Poacea sauvage (*Typha*) et une population évoluant dans des cultures de maïs avait également révélé des différences d'expression d'OBPs, de CSPs et d'ORs (Glaser et al., 2015).

Des gènes codant des heat-shock protéines (HSPs 40, 70 et 90), des protéines cuticulaires et des cytochromes P450 ont également été détectés comme différentiellement sous-exprimés dans les échantillons d'élevage par rapport aux sauvages. Les HSPs sont des protéines chaperonnes qui peuvent être présentes dans les cellules dans des conditions tout à fait normales. Elles ont été nommées ainsi du fait de la régulation transcriptionnelle positive des gènes codant ces protéines lors d'un choc thermique. Les HSPs sont aussi appelées des protéines de stress car elles répondent également au froid en activant plusieurs voies en cascade chez la drosophile (Burton et al., 1988). Elles sont également surexprimées en réponse à la déshydratation chez les moustiques *Aedes aegypti*, *Anopheles gambiae* et *Culex pipiens* (Benoit et al., 2010). Les HSPs 40, 70 et 90 sont également impliquées dans la réponse à la déshydratation chez la cécidomyie (moucheron) *Belgica antarctica* (Teets et al., 2012) et chez la cécidomyie *Polypedilum vanderplanki* (Cornette et al., 2010). De même, la cuticule joue le rôle de barrière à la perte d'eau et la transcription des gènes codant les protéines qui la compose est impactée par la déshydratation (Rajpurohit et al., 2013). Les cytochrome P450 sont impliqués dans la désintoxication par catalyse de diverses réactions, en particulier par mono-oxygénation (Nkya et al., 2013; Scott, 1999). Notre étude révèle que des HSPs, des protéines cuticulaires et une cytochrome P450 sont surexprimées dans les échantillons sauvages par rapport aux échantillons d'élevage. Cette observation est à mettre en relation avec les conditions d'élevage plus stables (donc moins stressantes) et optimales (y compris pour l'humidité) que celles rencontrées sur le terrain.

Conclusion

Notre étude transcriptomique comparative chez deux espèces sœurs de punaises hématophages vectrices de la maladie de Chagas, *R. prolixus* et *R. robustus*, a révélé un nombre de transcrits chimiosensoriels communs très important et une proximité très forte de l'expression des gènes entre ces deux espèces en conditions d'élevage. Cependant, l'environnement (élevage / sauvage) et le sexe impactent l'expression de nombreux gènes dont des gènes chimiosensoriels, des gènes impliqués dans la détoxification ou la résistance à la déshydratation. Par ailleurs, pour un même gène, l'isoforme la plus représentative diffère en fonction de l'espèce, du tissu, du sexe ou de l'environnement d'origine des punaises. Cependant, les observations issues de ce travail sont à confirmer par des analyses complémentaires. En effet, l'absence d'expression de certains gènes chimiosensoriels liée au sexe, à l'espèce ou à l'environnement serait à vérifier par des méthodes de PCR car une très faible expression de ces

gènes pourrait aboutir à leur absence de détection par les méthodes RNAseq. Par ailleurs, les échantillons d'élevage comprennent beaucoup plus d'individus par échantillon (18 à 21) que les échantillons sauvages (2 à 4) du fait du faible nombre collectés sur le terrain. Cependant, dans ces pools d'individus d'élevage même s'ils regroupent un plus grand nombre d'individus, il y a probablement moins de polymorphisme et donc potentiellement moins de variabilité d'expression que dans des pools d'individus de terrain. L'échantillon RMG constitué à partir de 4 individus s'écarte fortement de l'expression globale des autres échantillons. Cet écart, impliquant une augmentation de la variation entre nos différents échantillons, a probablement pour effet de diminuer la puissance de détection de gènes différentiellement exprimés. En contrepartie, cette perte de puissance renforce la confiance que nous pouvons avoir dans les gènes DE mis en évidence.

Le fait que l'expression des échantillons *R. robustus* d'élevage se rapproche davantage de celle de *R. prolixus* que de celle des *R. robustus* sauvages suggère soit une plasticité de l'expression des gènes de *R. robustus* en réponse à des modifications environnementales, soit une rapide évolution génétique de *R. robustus* au cours des différentes générations élevées en laboratoire en réponse aux pressions de sélection exercées par ce nouvel environnement. Dans les deux cas de figure, ceci suggère que *R. robustus* présente des potentialités à s'adapter aux anthroposystèmes. Cependant, l'analyse phylogénétique des *Rhodnius* que je présenterai au chapitre suivant révèle que la souche de *R. robustus* d'élevage se rapproche davantage de *R. prolixus* que de *R. robustus* de Guyane. Il est donc possible que la différence d'expression révélée entre les échantillons d'élevage et sauvages soit due au moins en partie à la divergence génétique qui sépare les individus de Guyane des individus d'élevage et non à une divergence environnementale. Par ailleurs, les analyses des isoformes ont montré des divergences notables entre les deux espèces. Si les différents épissages alternatifs ou les polymorphismes des gènes chimiosensoriels atteignent le site actif de la protéine, la fonction des isoformes pourrait en être modifiée. Dans ce cas, la capacité de *R. prolixus* à se domicilier contrairement à *R. robustus* pourrait provenir des fonctions acquises par la mise en place de ces isoformes. Ainsi, une étude sur les diversités fonctionnelles imputées aux variations d'isoformes détectées entre les deux espèces pourrait apporter d'autres éléments de réponse sur leur différence de capacité à coloniser les milieux anthropiques. Par ailleurs, nous pourrions également rechercher des traces de sélection éventuelles impliquant les quelques gènes chimiosensoriels présents chez *R. prolixus* et absents chez *R. robustus* afin de déterminer si ces gènes auraient un rôle dans l'adaptation de *R. prolixus* aux environnements domiciliés.

*Partie II : Analyse des transcriptomes chimiosensoriels chez différentes
espèces du genre Rhodnius*

Annexes

Contig	Gène	Blast NR ID	Annotation	%ident	Score	Evalue
TCONS_00177812	RPRC012014	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	39.55	181	2,00E-46
TCONS_00091786	RPRC005773	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	37.29	124	9,00E-28
TCONS_00086633	RPRC005347	gi 817196277	PREDICTED: protein takeout-like [Orussus abietinus]	37.68	149	2,00E-40
TCONS_00177776	RPRC011983	gi 242022241	protein takeout precursor, putative [Pediculus humanus corporis]humanus corporis]	37.50	92.0	5,00E-17
TCONS_00091778	RPRC010201	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	52.25	249	6,00E-70
TCONS_00043555	RPRC002529	gi 156546390	PREDICTED: protein takeout-like [Nasonia vitripennis]	23.64	75.1	3,00E-11
TCONS_00087871	RPRC009613	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	36.41	142	4,00E-33
TCONS_00114514	RPRC008440	gi 242009643	protein takeout precursor, putative [Pediculus humanus corporis]	50.34	142	4,00E-34
TCONS_00114299	RPRC002517	gi 662213264	PREDICTED: protein takeout-like [Diaphorina citri]	39.39	65.9	1,00E-08
TCONS_00091598	RPRC010085	gi 242024397	protein takeout precursor, putative [Pediculus humanus corporis]humanus corporis]	31.71	120	6,00E-27
TCONS_00043551	RPRC002524	gi 557757526	PREDICTED: protein takeout [Musca domestica]	27.74	80.9	9,00E-13
TCONS_00114302	RPRC002554	gi 328712782	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	59.38	242	1,00E-67
TCONS_00091589	RPRC010096	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	38.46	177	1,00E-45
TCONS_00114510	RPRC008451	gi 817065313	PREDICTED: protein takeout-like [Athalia rosae]	40.44	101	1,00E-20
TCONS_00143435	RPRC012883	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	31.40	59.3	3,00E-07
TCONS_00114408	RPRC002531	gi 389610927	takeout/JHBP like protein [Papilio polytes]	48.91	149	1,00E-35
TCONS_00070532	RPRC008276	gi 328717644	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	26.75	78.6	2,00E-12
TCONS_00091603	RPRC005775	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	44.88	194	5,00E-51
TCONS_00091792	RPRC005893	gi 18140056	takeout [Aedes aegypti]	29.95	98.6	6,00E-19
TCONS_00091622	RPRC005776	gi 646720336	Protein takeout [Zootermopsis nevadensis]	44.25	106	2,00E-21
TCONS_00091774	RPRC010202	gi 662206712	PREDICTED: protein takeout-like [Diaphorina citri]	64.29	314	2,00E-94
TCONS_00114300	RPRC002506	gi 189237515	PREDICTED: protein takeout isoform X1 [Tribolium castaneum]	58.68	320	5,00E-106
TCONS_00117821	RPRC006285	gi 646707272	Protein takeout [Zootermopsis nevadensis]	58.33	302	2,00E-93
TCONS_00091780	RPRC005774	gi 193636528	PREDICTED: protein takeout-like [Acyrtosiphon pisum]	50.84	191	8,00E-52
TCONS_00100030	RPRC009722	gi 211908973	70 kDa heat shock protein, partial [Pyrrhocoris apterus]	77.29	372	8,00E-125
TCONS_00061888	RPRC013236	gi 501294808	heat shock protein, putative [Riptortus pedestris]	92.38	201	1,00E-62
TCONS_00119846	RPRC014677	gi 399894429	heat shock protein 90 [Paratlanticus ussuriensis]	86.70	1046	0.0
TCONS_00175885	RPRC005100	gi 328785915	PREDICTED: heat shock factor protein isoform X4 [Apis mellifera]	36.07	97.4	4,00E-19
TCONS_00061948	RPRC013237	gi 699979714	heat shock protein, partial [Phenacoccus solenopsis]	80.86	495	1,00E-162
TCONS_00152935	RPRC002746	gi 820840593	PREDICTED: 97 kDa heat shock protein isoform X2 [Apis florea]	81.65	186	2,00E-51
TCONS_00024564	RPRC013971	gi 746259810	heat shock chaperone IbpB [Pectobacterium betavascularum]	80.13	248	3,00E-81
TCONS_00064939	RPRC013996	gi 85816372	heat shock protein 40 [Locusta migratoria]	61.16	394	8,00E-132
TCONS_00170974	XLOC_124125	gi 662222718	PREDICTED: heat shock 70 kDa protein cognate 1-like [Diaphorina citri]	96.61	122	2,00E-28

Partie II : Analyse des transcriptomes chimiosensoriels chez différentes espèces du genre Rhodnius

TCONS_00069056	RPRC009389	gi 85816372	heat shock protein 40 [Locusta migratoria]	73.77	290	9,00E-93
TCONS_00056392	RPRC013862	gi 746835673	PREDICTED: activator of 90 kDa heat shock protein ATPase homolog 1 isoform X2 [Acromyrmex echinator]	39.92	182	1,00E-46
TCONS_00149534	RPRC003899	gi 506954280	heat shock protein 90 [Apolygus lucorum]	93.47	807	0.0
TCONS_00151618	RPRC010472	gi 357614984	cuticular protein RR-2 motif 79 [Danaus plexippus]	85.71	136	8,00E-34
TCONS_00170437	RPRC011151	gi 357623617	cuticular protein RR-1 motif 32 [Danaus plexippus]	62.90	82.8	1,00E-17
TCONS_00120243	RPRC006716	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	75.76	107	1,00E-26
TCONS_00170416	RPRC000052	gi 817222440	PREDICTED: endocuticle structural glycoprotein SgAbd-8-like [Orussus abietinus]	68.29	110	2,00E-23
TCONS_00012845	RPRC005240	gi 290560901	cuticular protein RR-2 motif 111 precursor [Bombyx mori]	73.13	102	7,00E-22
TCONS_00066718	RPRC005931	gi 662225759	PREDICTED: larval cuticle protein 1-like, partial [Diaphorina citri]	76.00	127	4,00E-30
TCONS_00037783	RPRC006627	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	62.12	84.3	1,00E-17
TCONS_00170424	RPRC011145	gi 357619654	cuticular protein RR-1 motif 8 [Danaus plexippus]	62.24	105	4,00E-21
TCONS_00125494	RPRC005396	gi 289684243	cuticular protein RR-2 family member 7 precursor [Nasonia vitripennis]	76.81	114	1,00E-24
TCONS_00151588	RPRC010681	gi 768407863	PREDICTED: cuticle protein 8-like [Plutella xylostella]	80.82	129	6,00E-30
TCONS_00170425	RPRC011152	gi 357619654	cuticular protein RR-1 motif 8 [Danaus plexippus]	62.50	99.4	1,00E-21
TCONS_00170435	RPRC011143	gi 642924550	PREDICTED: endocuticle structural glycoprotein SgAbd-8 isoform X2 [Tribolium castaneum]	72.17	147	3,00E-38
TCONS_00170436	XLOC_123803	gi 646704638	Endocuticle structural glycoprotein SgAbd-2, partial [Zootermopsis nevadensis]	40.56	121	1,00E-26
TCONS_00170426	RPRC011144	gi 357619654	cuticular protein RR-1 motif 8 [Danaus plexippus]	67.02	107	3,00E-22
TCONS_00125488	RPRC005398	gi 289684241	cuticular protein RR-2 family member 4 precursor [Nasonia vitripennis]	80.88	115	6,00E-28
TCONS_00069210	RPRC009907	gi 817222440	PREDICTED: endocuticle structural glycoprotein SgAbd-8-like [Orussus abietinus]	62.96	118	5,00E-29
TCONS_00120208	RPRC006619	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	70.59	107	2,00E-26
TCONS_00116859	RPRC001565	gi 389609701	cuticular protein PxutCPR147 [Papilio xuthus]	57.63	74.7	6,00E-12
TCONS_00086405	RPRC001297	gi 290563239	cuticular protein RR-1 motif 47 precursor [Bombyx mori]	49.82	246	1,00E-68
TCONS_00151264	RPRC010684	gi 768407863	PREDICTED: cuticle protein 8-like [Plutella xylostella]	80.82	129	4,00E-35
TCONS_00153442	RPRC013399	gi 646699612	Endocuticle structural glycoprotein SgAbd-2 [Zootermopsis nevadensis]	58.23	101	1,00E-24
TCONS_00131264	RPRC002431	gi 817214818	PREDICTED: cuticle protein 21-like [Orussus abietinus]	73.68	119	9,00E-30
TCONS_00120249	RPRC006617	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	69.86	105	2,00E-22
TCONS_00170438	RPRC001982	gi 568250246	pupal cuticle protein 78E [Anopheles darlingi]	61.45	115	4,00E-29
TCONS_00120251	XLOC_088575	gi 665800845	PREDICTED: adult-specific cuticular protein ACP-22-like [Microplitis demolitor]	71.43	108	3,00E-23
TCONS_00116854	RPRC000036	gi 768407863	PREDICTED: cuticle protein 8-like [Plutella xylostella]	77.92	127	2,00E-29
TCONS_00120238	RPRC000550	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	68.83	110	5,00E-27
TCONS_00012843	RPRC005241	gi 290560901	cuticular protein RR-2 motif 111 precursor [Bombyx mori]	73.13	102	4,00E-24
TCONS_00081247	RPRC004265	gi 501291136	cuticle protein, putative [Riptortus pedestris]	95.59	136	1,00E-35
TCONS_00131266	RPRC002446	gi 766931077	PREDICTED: larval cuticle protein A2B-like [Ceratosolen solmsi marchali]	77.14	120	2,00E-31
TCONS_00021983	RPRC000069	gi 501291136	cuticle protein, putative [Riptortus pedestris]	87.14	129	1,00E-29
TCONS_00036513	RPRC004479	gi 662197351	PREDICTED: endocuticle structural glycoprotein SgAbd-2-like [Diaphorina citri]	73.03	124	5,00E-33

*Partie II : Analyse des transcriptomes chimiosensoriels chez différentes
espèces du genre Rhodnius*

TCONS_00170429	RPRC001981	gi 345493131	PREDICTED: cuticle protein CP14.6 [Nasonia vitripennis]	58.24	99.4	5,00E-20
TCONS_00115498	XLOC_085406	gi 357607580	TPAputative cuticle protein [Danaus plexippus]	68.29	118	5,00E-26
TCONS_00125739	RPRC013838	gi 289684241	cuticular protein RR-2 family member 4 precursor [Nasonia vitripennis]	79.41	116	1,00E-24
TCONS_00116064	RPRC013064	gi 817214821	PREDICTED: cuticle protein 21-like [Orussus abietinus]	89.86	134	4,00E-31
TCONS_00116059	RPRC013133	gi 766931077	PREDICTED: larval cuticle protein A2B-like [Ceratosolen solmsi marchali]	83.33	129	3,00E-30
TCONS_00131292	RPRC002432	gi 501291136	cuticle protein, putative [Riptortus pedestris]	75.36	114	9,00E-28
TCONS_00181845	XLOC_132458	gi 290560618	cuticular protein RR-2 motif 124 precursor [Bombyx mori]	63.38	87.4	6,00E-18
TCONS_00109618	RPRC001502	gi 389608381	cuticular protein PxutCPR10 [Papilio xuthus]	62.83	144	9,00E-38
TCONS_00011303	XLOC_009783	gi 642922070	PREDICTED: cuticle protein 19-like [Tribolium castaneum]	74.07	84.0	1,00E-17
TCONS_00131308	RPRC002429	gi 389611355	cuticular protein PpolCPR67A [Papilio polytes]	81.01	136	2,00E-35
TCONS_00054988	RPRC009528	gi 645013148	PREDICTED: cuticle protein 19 [Nasonia vitripennis]	89.23	122	7,00E-32
TCONS_00115264	RPRC009820	gi 662200828	PREDICTED: larval cuticle protein A3A-like [Diaphorina citri]	59.18	68.2	4,00E-10
TCONS_00021984	RPRC006190	gi 290563378	cuticular protein RR-2 motif 135 precursor [Bombyx mori]	49.47	91.7	4,00E-16
TCONS_00131285	RPRC002433	gi 766931077	PREDICTED: larval cuticle protein A2B-like [Ceratosolen solmsi marchali]	81.43	127	4,00E-32
TCONS_00131288	RPRC002448	gi 290560754	cuticular protein RR-2 motif 73 precursor [Bombyx mori]	71.64	106	2,00E-25
TCONS_00131294	RPRC000173	gi 817214821	PREDICTED: cuticle protein 21-like [Orussus abietinus]	74.19	99.8	2,00E-22
TCONS_00071934	RPRC004403	gi 240848893	cuticular protein CPG12-like precursor [Acyrtosiphon pisum]	82.43	76.3	4,00E-11
TCONS_00012340	RPRC006463	gi 242015941	Endocuticle structural glycoprotein SgAbd-4, putative [Pediculus humanus corporis]	70.37	139	1,00E-38
TCONS_00125737	RPRC005594	gi 289684243	cuticular protein RR-2 family member 7 precursor [Nasonia vitripennis]	71.23	113	2,00E-27
TCONS_00120343	XLOC_088660	gi 354549517	cuticular proteinRR-2 motif 126 [Antheraea yamamai]	68.00	106	3,00E-22
TCONS_00163014	RPRC006462	gi 1706197	RecName: Full=Pupal cuticle protein C1B; AltName: Full=TM-PCP C1B; Short=TM-C1B [Tenebrio molitor]	82.22	83.2	2,00E-14
TCONS_00125490	XLOC_092083	gi 289684241	cuticular protein RR-2 family member 4 precursor [Nasonia vitripennis]	41.18	70.9	9,00E-10
TCONS_00131316	XLOC_096195	gi 817187373	PREDICTED: cuticle protein 21-like [Orussus abietinus]	82.35	125	9,00E-28
TCONS_00037782	RPRC006628	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	75.76	106	2,00E-26
TCONS_00175401	RPRC002929	gi 817181520	PREDICTED: endocuticle structural glycoprotein SgAbd-2-like [Orussus abietinus]	52.38	110	7,00E-28
TCONS_00042337	RPRC012770	gi 662184166	PREDICTED: endocuticle structural glycoprotein SgAbd-9-like [Diaphorina citri]	55.34	102	3,00E-21
TCONS_00011505	RPRC010438	gi 389609703	cuticular protein PxutCPR107 [Papilio xuthus]	75.76	108	1,00E-26
TCONS_00042348	RPRC000280	gi 357623612	TPAputative cuticle protein [Danaus plexippus]	39.51	60.8	2,00E-07
TCONS_00163041	XLOC_118490	gi 1706197	RecName: Full=Pupal cuticle protein C1B; AltName: Full=TM-PCP C1B; Short=TM-C1B [Tenebrio molitor]	82.22	82.8	3,00E-13
TCONS_00066735	RPRC005844	gi 478859275	cuticular protein BJ [Laodelphax striatella]	79.45	117	1,00E-24
TCONS_00127638	XLOC_093439	gi 646680486	Naked cuticle-like protein [Zootermopsis nevadensis]	82.05	68.9	2,00E-08
TCONS_00045755	RPRC012706	gi 328718464	PREDICTED: cuticle protein 16.5, isoform B-like [Acyrtosiphon pisum]	75.51	77.4	1,00E-15
TCONS_00127626	RPRC005135	gi 646680486	Naked cuticle-like protein [Zootermopsis nevadensis]	82.05	68.9	2,00E-08
TCONS_00131295	RPRC002430	gi 766931077	PREDICTED: larval cuticle protein A2B-like [Ceratosolen	78.69	107	4,00E-23

*Partie II : Analyse des transcriptomes chimiosensoriels chez différentes espèces du genre *Rhodnius**

			solmsi marchali]			
TCONS_00051661	RPRC005300	gi 242014917	structural constituent of cuticle, putative [Pediculus humanus corporis]	38.24	110	6,00E-22
TCONS_00131307	XLOC_096190	gi 290560754	cuticular protein RR-2 motif 73 precursor [Bombyx mori]<>TPA: putative cuticle protein [Bombyx mori]	66.32	123	2,00E-29
TCONS_00066733	RPRC005928	gi 268607717	cuticular protein precursor [Tribolium castaneum]	74.79	180	5,00E-45
TCONS_00062923	RPRC001110	gi 289684241	cuticular protein RR-2 family member 4 precursor [Nasonia vitripennis]	47.56	84.0	1,00E-13
TCONS_00126001	XLOC_092368	gi 289684243	cuticular protein RR-2 family member 7 precursor [Nasonia vitripennis]	75.36	113	1,00E-25
TCONS_00163092	XLOC_118540	gi 1706197	RecName: Full=Pupal cuticle protein C1B; AltName: Full=TM-PCP C1B; Short=TM-C1B [Tenebrio molitor]	82.22	82.8	7,00E-16

Tableau 1: annotation par blast sur NR (version mars 2015) des transcrits codant pour des protéines takeout, pour des protéines heat shock ou des protéines cuticulaires. Résultats de blasts ont été décrits: pour tous les contigs qui correspondaient avec une protéine d'intérêt de la base de données non redondante des protéines, l'ID de cette protéine (Blast NR ID) et la dénomination (Annotation) ont été indiqués. Pour chaque match, nous avons indiqué le pourcentage d'identité (% ident), le score et la Evalue.

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
CSP1	RPRC013223	TCONS_00061919	1,150	1,150	12,912	12,574	15056,800	15056,800	4599,770	4599,770	1,082	0,630	16,901	16,901	0,998	0,998	24,979	24,301
CSP2	RPRC013221	TCONS_00061922	0,164	0,164	12,565	12,565	22453,600	22453,600	5945,730	5945,730	5,950	5,950	10,913	10,913	0,399	0,399	24,496	24,496
CSP3	RPRC013280	TCONS_00061923	63,425	63,425	258,789	69,203	99482,500	1115,180	22673,800	452,148	289,711	77,814	97,625	56,739	59,108	57,006	462,787	87,817
CSP4	RPRC013285	TCONS_00061926	0,164	0,164	2,540	2,540	1603,480	1603,480	70,238	70,238	3,246	3,246	0,250	0,250	0,188	0,188	4,868	4,868
CSP5	RPRC013218	TCONS_00061929	22,094	3,939	100,863	64,313	27430,300	222,376	5115,940	87,574	81,459	79,180	67,755	9,196	27,622	14,418	168,576	103,729
CSP6	RPRC000332	TCONS_00061932	57,825	57,825	78,710	74,854	316,369	103,987	68,671	31,244	109,200	105,275	34,296	33,081	42,936	42,024	129,373	123,485
CSP7	RPRC010107	TCONS_00095976	293,395	70,884	669,521	326,670	518,771	14,449	3119,870	482,548	856,829	415,863	231,463	66,953	276,800	96,073	1078,840	532,078
CSP8	RPRC000631	TCONS_00002593	51,991	51,991	4,440	4,309	121,626	34,881	111,074	0,000	31,267	31,092	9,314	9,314	30,030	29,899	0,811	0,811
CSP9	RPRC000470	TCONS_00093752	53,160	51,847	66,833	58,745	542,151	0,000	891,196	49,923	92,301	86,989	32,250	25,902	40,534	39,332	105,759	90,671
CSP10	RPRC007979	TCONS_00171934	137,727	134,568	164,968	164,968	49,643	45,765	833,055	421,016	102,367	100,262	209,408	209,408	140,886	139,306	185,892	185,892
CSP11	RPRC011935	TCONS_00177364	2,223	2,223	307,830	307,830	2000,620	2000,620	19790,800	19790,800	3,208	3,208	408,715	408,715	1,746	1,746	614,392	614,392
CSP12	RPRC011928	TCONS_00177373	13,529	13,529	6,744	6,744	7585,480	7585,480	4064,000	4064,000	12,458	12,458	5,553	5,553	7,504	7,504	12,007	12,007
CSP13	RPRC011926	TCONS_00177804	35,142	17,858	14,618	1,056	938,256	585,251	766,084	496,908	22,622	9,398	20,297	3,916	22,858	9,985	18,662	0,000
CSP14	RPRC006791	TCONS_00099137	0,164	0,164	11,450	11,450	0,000	0,000	0,000	0,000	1,304	1,304	14,072	14,072	0,989	0,989	21,086	21,086
CSP15	RPRC000546	TCONS_00025500	1,339	1,339	0,106	0,106	0,000	0,000	0,000	0,000	0,783	0,783	0,250	0,250	0,775	0,775	0,000	0,000
CSP16	RPRC000327	TCONS_00150597	68,124	16,320	84,198	73,819	104,251	0,000	504,733	30,473	120,583	94,215	37,096	15,091	109,740	73,463	17,040	17,033
CSP17	RPRC007134	TCONS_00000503	0,556	0,556	0,634	0,634	484,021	484,021	416,528	416,528	0,261	0,261	0,955	0,955	0,912	0,912	0,000	0,000
CSP18	RPRC000242	TCONS_00077713	5,722	5,722	34,262	34,262	0,000	0,000	0,000	0,000	4,233	4,233	45,264	45,264	4,764	4,764	64,717	64,717
CSP19	RPRC000126	TCONS_00078853	0,000	0,000	0,923	0,923	253,180	253,180	66,971	66,971	0,666	0,666	0,564	0,564	0,923	0,923	0,000	0,000
OBP1	RPRC010284	TCONS_00154195	37,428	0,392	34,301	22,014	605,647	605,647	42,470	42,470	55,257	29,472	15,430	0,141	23,399	0,302	59,233	43,816
OBP2	RPRC000378	TCONS_00154188	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
OBP3	RPRC000527	TCONS_00163866	201,856	84,101	239,964	18,040	27,304	2,567	93,106	1,688	208,231	58,856	246,291	21,264	166,749	53,050	348,285	14,081
OBP4	RPRC000274	TCONS_00054965	53,800	29,206	15,270	9,386	375,831	49,093	259,606	47,189	14,688	2,737	41,539	29,248	39,272	21,936	5,797	4,106
OBP5	XLOC_059757	TCONS_00077871	668,395	491,555	1227,110	884,293	541,111	518,436	133,942	129,823	1690,740	1350,040	391,004	156,716	607,783	374,888	1907,040	1510,360
OBP6	RPRC004413	TCONS_00168603	109,901	58,542	194,463	8,260	915,917	155,643	5769,310	2074,120	240,755	17,558	91,796	32,483	193,427	32,667	111,974	9,727
OBP7	RPRC000311	TCONS_00168607	0,164	0,164	0,211	0,211	14,893	14,893	60,437	60,437	0,000	0,000	0,391	0,391	0,294	0,294	0,000	0,000
OBP8	RPRC000564	TCONS_00168610	5,546	5,538	15,155	15,147	682,594	0,000	156,810	0,000	14,046	14,035	9,859	9,853	5,839	5,835	24,178	24,162

OBP9	RPRC000194	TCONS_00168611	0,392	0,392	0,000	0,000	0,000	0,000	0,000	0,000	0,261	0,261	0,000	0,000	0,196	0,196	0,000	0,000
OBP10	RPRC000560	TCONS_00168612	194,590	194,590	22,786	22,786	11,920	11,920	1,819	1,819	130,155	130,155	29,953	29,953	113,799	113,799	12,563	12,563
OBP11	RPRC004408	TCONS_00168667	182,885	91,827	316,326	217,538	9593,550	181,873	5908,160	51,494	358,374	218,581	185,316	132,688	242,437	175,786	330,661	175,331
OBP12	RPRC000492	TCONS_00168613	49,706	49,706	349,358	347,412	1529,010	333,730	4436,430	1180,540	108,619	108,619	390,329	387,734	68,869	68,869	610,684	606,792
OBP13	RPRC000177	TCONS_00168615	127,670	127,670	56,003	56,003	84,393	53,896	594,573	290,946	85,571	85,571	74,213	74,213	114,646	114,646	10,384	10,384
OBP14	RPRC004407	TCONS_00168618	0,392	0,392	10,813	10,813	3445,240	3445,240	3694,840	3694,840	6,752	6,752	7,926	7,926	0,302	0,302	21,415	21,415
OBP15	RPRC008413	TCONS_00067373	2,122	2,122	0,500	0,500	0,000	0,000	0,000	0,000	1,972	1,972	0,109	0,109	1,561	1,561	0,000	0,000
OBP16	RPRC011610	TCONS_00059698	41,008	20,418	7,132	4,698	914,455	867,846	524,259	499,757	23,936	11,348	12,912	8,527	26,013	13,284	3,246	3,246
OBP17	RPRC000118	TCONS_00177251	51,764	51,764	4,210	4,210	327,645	76,803	4320,450	1149,510	33,414	33,414	6,708	6,708	26,199	26,199	7,785	7,785
OBP18	RPRC001924	TCONS_00160295	1,478	1,478	2,687	0,740	2415,140	453,901	13025,100	3415,780	0,000	0,000	4,567	1,972	1,479	1,479	3,893	0,000
OBP19	RPRC000593	TCONS_00160301	52,434	38,433	21,035	11,718	52,125	6,188	52,270	0,000	36,873	22,725	26,130	18,522	47,252	30,935	0,000	0,000
OBP20	RPRC007008	TCONS_00160304	287,483	209,831	280,599	80,745	124056,000	630,965	32448,300	175,284	461,983	239,114	103,805	8,434	199,598	108,467	449,486	154,388
OBP21	RPRC001925	TCONS_00160312	132,132	16,637	133,489	76,311	342,538	16,258	1582,800	25,080	91,935	14,933	174,139	97,907	133,457	22,424	132,196	124,411
OBP22	RPRC000115	TCONS_00069219	23,495	12,700	1,692	0,000	394,663	0,000	743,216	0,000	12,269	8,467	5,650	0,000	13,439	6,350	0,000	0,000
OBP23	RPRC000174	TCONS_00069224	106,588	89,827	1402,770	1401,730	7247,910	285,541	3020,230	277,963	106,639	94,577	1834,780	1834,280	72,686	63,265	2766,750	2766,750
OBP24	RPRC009905	TCONS_00069231	37,520	29,719	62,193	58,677	3152,200	502,531	36858,100	7294,970	75,052	70,364	32,885	27,685	57,579	53,668	46,747	39,736
OBP25	RPRC000257	TCONS_00069263	28,093	28,093	24,883	24,883	15,038	15,038	96,863	96,863	26,182	26,182	25,724	25,724	24,573	24,573	28,714	28,714
OBP26	RPRC000182	TCONS_00069276	4,598	4,598	8,004	4,441	1099,600	0,000	339,756	0,000	4,327	0,000	9,410	8,987	7,057	6,740	6,491	0,000
OBP27	RPRC009914	TCONS_00069282	24,518	24,518	6,167	6,167	22,312	22,312	142,109	132,729	15,881	15,881	8,687	8,687	15,992	15,992	4,868	4,868
OBP28	RPRC009906	TCONS_00069213	9,222	0,000	45,611	0,000	37,232	17,375	53,904	42,470	47,526	0,000	19,437	0,000	11,439	0,000	77,566	0,000

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
Or1	RPRC000579	TCONS_00150419	266,205	76,392	172,392	65,437	2,482	0,000	47,370	2,345	231,113	95,544	176,214	42,634	258,608	96,561	93,773	14,144
Or3	RPRC000059	TCONS_00037174	63,416	63,416	45,429	45,429	0,000	0,000	13,068	13,068	36,568	36,568	66,281	66,281	60,133	60,133	34,007	34,007
Or5	RPRC000083	TCONS_00135190	75,688	69,379	19,857	2,588	0,000	0,000	0,000	0,000	65,637	45,423	11,297	4,280	55,754	37,277	3,893	0,000
Or10	RPRC000601	TCONS_00164670	5,546	0,328	102,839	79,306	0,000	0,000	3,267	3,267	129,265	105,460	11,551	0,501	7,108	0,376	197,007	158,190
Or12	RPRC000376	TCONS_00068718	20,603	20,603	3,188	3,188	0,000	0,000	1,633	0,000	2,794	2,794	15,193	15,193	13,490	13,490	0,000	0,000
Or13	RPRC000362	TCONS_00056495	43,846	37,383	95,263	90,882	255,662	132,670	39,203	3,374	128,044	124,579	28,204	21,519	115,969	108,845	2,434	1,457
Or14	RPRC000166	TCONS_00035558	11,003	11,003	53,801	53,801	2,482	2,482	16,334	16,334	64,205	64,205	14,865	14,865	20,032	20,032	78,542	78,542
Or15	RPRC000547	TCONS_00143424	148,358	148,358	95,665	95,665	37,232	12,556	26,135	4,956	150,221	150,221	76,238	76,238	93,670	93,670	152,348	152,348
Or16	RPRC000420	TCONS_00126179	20,124	20,124	259,544	259,375	0,000	0,000	4,900	2,209	330,044	329,818	29,431	29,431	257,435	257,266	24,342	24,342
Or18	RPRC000616	TCONS_00126117	168,358	168,358	29,807	29,807	0,000	0,000	0,000	0,000	117,897	117,897	34,085	34,085	98,416	98,416	31,140	31,140
Or19	RPRC000449	TCONS_00109227	79,721	76,344	62,254	60,308	7,446	7,445	9,801	9,799	86,354	84,103	49,798	47,203	95,299	93,611	13,630	9,737
Or20	RPRC000573	TCONS_00037649	2,135	2,135	3,744	1,798	14,893	14,893	19,601	6,834	0,000	0,000	6,415	3,820	2,865	2,865	3,893	0,000
Or21	RPRC006570	TCONS_00047083	0,000	0,000	0,211	0,211	0,000	0,000	0,000	0,000	0,000	0,000	0,282	0,282	0,211	0,211	0,000	0,000
Or22	RPRC000375	TCONS_00047096	0,000	0,000	48,628	48,628	0,000	0,000	0,000	0,000	4,868	4,868	59,969	59,969	0,000	0,000	96,834	96,834
Or23	RPRC000149	TCONS_00047098	115,926	55,561	169,866	133,294	0,000	0,000	1,632	1,632	160,290	145,736	143,482	69,030	175,335	108,580	104,988	104,988
Or25	RPRC000029	TCONS_00062915	393,266	77,525	1241,510	494,467	447,933	120,219	781,301	242,582	605,483	25,813	1312,050	685,159	388,696	79,219	2098,900	908,02
Or26	RPRC000619	TCONS_00062892	44,833	43,975	29,891	16,762	29,784	0,000	117,608	45,591	11,146	8,441	58,597	43,224	42,494	38,749	19,627	0,000
Or29	RPRC000125	TCONS_00019062	6,531	6,330	100,117	97,286	2,482	0,002	6,534	6,534	130,570	130,030	7,273	3,905	101,030	100,450	4,704	0,002
Or30	RPRC000209	TCONS_00040217	4,093	4,093	3,789	3,789	0,000	0,000	80,039	80,039	1,846	1,846	5,935	5,935	5,430	5,430	0,811	0,811
Or31	RPRC000423	TCONS_00069234	2,463	2,463	24,752	24,752	2,482	2,482	0,000	0,000	30,042	30,042	4,603	4,603	23,112	23,112	5,744	5,744
Or32	RPRC000043	TCONS_00059354	12,480	5,937	4,372	2,749	7,446	7,446	3,267	0,000	8,951	2,425	5,198	5,198	8,989	5,718	3,246	0,000
Or33	RPRC000235	TCONS_00059353	77,361	0,415	389,202	330,141	4,964	3,039	19,601	0,000	64,619	32,731	505,890	407,734	59,523	0,208	736,718	660,282
Or36	RPRC000337	TCONS_00036060	52,333	0,000	64,034	58,471	2,482	2,479	8,167	2,566	33,458	2,704	86,809	75,258	31,729	0,000	116,944	116,943
Or37	RPRC000456	TCONS_00036062	72,080	72,080	62,305	54,520	69,501	69,501	27,769	27,215	60,070	60,070	71,056	60,676	73,114	73,114	50,461	34,890
Or38	RPRC000530	TCONS_00171686	63,336	39,342	83,088	74,447	4,964	0,000	1,633	0,000	62,555	51,694	90,454	73,797	73,478	53,245	82,557	81,746
Or39	RPRC000041	TCONS_00121637	489,756	170,186	193,827	13,302	191,126	146,086	171,511	147,776	419,920	99,981	165,020	31,211	347,669	98,395	182,073	0,000

Or40	RPRC000229	TCONS_00034389	471,240	471,240	484,975	484,975	49,643	49,643	4,900	1,760	727,064	727,064	233,730	233,730	708,424	708,424	24,342	24,342
Or41	RPRC000146	TCONS_00081697	26,743	26,743	30,537	30,445	12,411	12,411	6,534	5,995	3,277	3,154	55,268	55,268	24,445	24,353	38,927	38,927
Or42	RPRC000108	TCONS_00006024	13,643	4,648	48,364	36,351	0,000	0,000	62,071	35,810	61,216	49,591	12,364	1,976	53,157	38,270	4,057	0,811
Or43	RPRC000140	TCONS_00086186	3,802	3,330	4,633	4,633	22,339	11,374	35,936	35,933	6,114	6,114	2,598	2,283	6,534	6,298	0,000	0,000
Or47	RPRC000413	TCONS_00086206	38,402	29,131	48,881	33,301	0,000	0,000	13,068	7,947	18,127	3,649	72,649	60,173	33,854	22,083	68,455	51,568
Or49	RPRC000433	TCONS_00125502	0,328	0,328	13,312	13,312	0,000	0,000	0,000	0,000	1,333	1,333	16,636	16,636	1,798	1,798	23,356	23,356
Or51	RPRC000507	TCONS_00079525	53,985	52,492	19,882	8,057	211,195	207,178	122,436	122,436	29,406	13,640	33,092	32,097	33,161	32,356	27,425	3,894
Or52	RPRC000201	TCONS_00072697	94,008	71,053	95,698	58,451	12,411	12,411	0,000	0,000	113,124	78,976	77,145	46,328	93,371	50,512	98,663	86,931
Or53	RPRC002122	TCONS_00147337	211,316	137,971	96,865	22616,0	0,000	0,000	52,270	0,000	192,571	100,169	77,458	8,074	191,327	81,182	22,391	0,000
Or54	RPRC000055	TCONS_00147331	10,863	9,519	11,680	1354626,0	0,000	0,000	4,900	0,000	14,957	14,173	7,859	7,587	11,026	10,235	12,171	12,171
Or57	RPRC000318	TCONS_00086863	32,489	1,402	864,707	833,968	0,000	0,000	3,267	3,267	23,758	53328,0	1150,840	1111,690	18,888	0,832	1724,130	1667,670
Or58	RPRC000526	TCONS_00035788	110,008	79,541	28,677	10,926	14,893	2,661	44,103	0,000	75,980	51,829	35,595	15,766	77,760	49,479	11,842	2,434
Or59	RPRC004587	TCONS_00097321	21,892	21,892	15,258	15,016	24,822	3,176	1,633	1,633	8,553	8,230	26,386	26,386	20,283	20,041	11,842	11,842
Or60	RPRC002105	TCONS_00088042	0,000	0,000	294,563	145,780	0,000	0,000	0,000	0,000	299,096	109,178	93,655	85,196	0,174	0,000	588,779	291,561
Or61	RPRC000208	TCONS_00088071	55,733	55,240	60,860	60,408	0,000	0,000	26,135	0,000	78,437	77,975	39,865	39,396	67,599	66,900	42,256	42,256
Or64	RPRC000095	TCONS_00088063	0,947	0,947	0,634	0,634	37,232	37,232	0,000	0,000	0,522	0,522	0,955	0,955	1,108	1,108	0,000	0,000
Or65	RPRC000132	TCONS_00088060	567,553	361,536	207,461	136,906	57,090	0,002	3,267	3,250	431,330	279,891	223,654	143,674	421,467	254,172	139,541	127,003
Or66	RPRC000452	TCONS_00088055	28,397	21,583	53,501	37,494	4,964	0,000	1,633	0,000	31,059	8,274	59,207	56,107	53,094	46,648	29,211	3,276
Or67	RPRC000222	TCONS_00088052	92,957	92,957	106,090	104,143	4,964	3,444	3,267	2,241	165,597	165,597	37,827	35,232	148,999	148,999	7,138	3,246
Or68	RPRC000541	TCONS_00088049	104,378	104,378	122,699	54,589	9,929	0,000	0,000	0,000	177,123	86,310	56,061	56,061	171,643	105,347	6,491	2,862
Or70	RPRC000459	TCONS_00088088	2,552	2,552	5,353	5,353	0,000	0,000	1,633	1,633	5,304	5,304	3,534	3,534	2,654	2,654	7,950	7,950
Or71	RPRC000490	TCONS_00088086	65,890	59,167	110,806	105,375	4,964	0,000	0,000	0,000	106,332	105,596	85,335	74,348	114,797	106,005	57,907	57,907
Or72	RPRC000053	TCONS_00088084	0,392	0,392	0,000	0,000	24,187	24,187	3,267	3,267	0,261	0,261	0,000	0,000	0,196	0,196	0,000	0,000
Or73	RPRC000585	TCONS_00088082	129,514	71,675	108,189	46,518	10,563	0,000	3,267	0,001	167,235	62,078	63,359	47,729	144,547	53,956	56,798	56,798
Or74	RPRC000200	TCONS_00088092	444,168	34,502	1173,010	1016,750	0,000	0,000	10,051	0,000	398,466	75,794	1461,650	1302,870	402,092	56,388	1986,000	1955,220
Or75	RPRC000245	TCONS_00088047	0,000	0,000	57,756	57,756	2,482	2,482	6,284	6,284	75,496	75,496	1,513	1,513	50,487	50,487	14,539	14,539
Or77	RPRC000468	TCONS_00161181	472,772	467,571	545,525	313,940	17,375	3,849	101,273	2,089	710,230	494,874	332,317	235,427	622,183	459,766	319,456	175,919
Or79	RPRC000202	TCONS_00118300	195,853	189,697	51,206	35,540	7,446	0,002	16,334	16,334	146,795	134,353	52,049	39,499	133,963	124,550	30,340	11,678
Or80	RPRC000441	TCONS_00118296	47,156	0,000	110,723	70,444	4,964	2,473	0,000	0,000	136,815	92,961	42,253	0,965	37,420	0,723	193,761	139,442
Or81	RPRC000363	TCONS_00118293	38,238	36,439	9,927	7,261	2,482	0,000	0,000	0,000	10,640	6,266	28,088	27,709	28,235	24,670	1,623	1,623

Or82	RPRC002113	TCONS_00118290	169,542	17,543	333,551	279,914	9,929	0,000	3,267	1,633	498,735	366,940	59,027	17,974	110,882	11,534	614,879	554,303
Or83	RPRC006981	TCONS_00036175	430,414	188,670	604,347	348,798	32,268	0,000	31,035	0,000	760,588	578,906	332,150	2729033,0	411,773	175,559	815,561	535,147
Or84	RPRC000431	TCONS_00036166	45,768	60,015	89,273	60,015	4,964	0,000	73,505	0,000	54,959	17,638	94,583	70,968	98,851	61,073	26,612	10,762
Or86	RPRC000079	TCONS_00036159	5,950	5,950	344,919	344,919	0,000	0,000	4,217	4,217	4,958	4,958	458.9	458,900	5,342	5,342	685,104	685,104
Or87	RPRC000317	TCONS_00036157	175,934	59,748	128,488	74,102	2,482	0,000	21,918	14,765	142,475	48,304	146,132	90,331	162,268	75,344	108,373	57,263
Or88	RPRC000607	TCONS_00032721	199,027	199,027	73,538	73,538	7,446	6,050	0,000	0,000	131,370	131,370	99,366	99,366	132,975	132,975	80,154	80,154
Or89	RPRC000165	TCONS_00174276	0,141	0,141	0,000	0,000	0,000	0,000	0,000	0,000	0,094	0,094	0,000	0,000	0,071	0,071	0,000	0,000
Or90	RPRC000249	TCONS_00174272	595,282	422,165	132,056	40,633	2,482	2,482	9,801	2,566	288,865	209,875	284,065	125,746	398,392	247,823	62,609	7,785
Or91	RPRC000093	TCONS_00174269	80,604	58,743	86,034	52,086	0,000	0,000	0,000	0,000	49,572	31,055	118,876	77,556	74,109	38,639	104,455	85,638
Or97	RPRC002120	TCONS_00143678	30,797	25,754	43,910	30,154	0,000	0,000	6,534	0,000	57,590	46,373	21,488	11,002	30,586	14,724	57,445	56,615
Or99	RPRC000065	TCONS_00143672	252,229	148,835	142,026	124,565	14,893	0,000	31,035	0,000	294,301	235,852	63,220	29,458	246,315	180,726	43,651	36,513
Or100	RPRC000522	TCONS_00143666	37,365	14,123	175,829	151,993	12,411	12,410	29,402	24,686	179,616	157,572	79,733	54,500	70,876	44,630	247,272	228,848
Or101	RPRC000188	TCONS_00028771	48,760	6,889	177,914	104,096	2,482	0,000	8,167	0,000	232,351	138,795	37,375	4,593	44,476	3,445	315,636	208,192
Or104	RPRC009806	TCONS_00075071	37,998	37,439	7,280	3,752	0,000	0,000	1,633	1,633	11,820	7,009	23,218	22,953	23,033	22,472	6,491	0,000
Or105	RPRC000255	TCONS_00121651	185,337	0,195	4000,520	3772,420	1027,610	0,000	1051,940	0,000	281,022	29,546	5176,560	5000,48	187,820	0,101	7810,740	7544,83
Or106	RPRC000371	TCONS_00139335	205,116	60,225	520,581	423,715	17,375	14,893	27,769	0,401	199,613	75,475	631,240	529,628	220,804	73,136	804,670	761,383
Or108	RPRC000301	TCONS_00139332	255,496	255,496	55,967	32,612	19,857	8,400	1,633	0,000	206,000	206,000	38,954	7,813	131,554	131,554	104,321	57,610
Or112	RPRC000555	TCONS_00120491	68,063	15,490	40,856	19,911	168,787	53,479	119,241	85,025	37,790	27,921	62,060	8,954	51,603	12,598	46,568	30,116
Or114	RPRC004306	TCONS_00087779	7,769	7,406	5,978	5,651	2,482	2,482	0,000	0,000	6,435	6,314	6,715	6,158	6,211	5,703	7,303	7,303
OrCo	RPRC000476	TCONS_00139686	5,874	0,000	18,211	0,000	173,595	160,786	73,076	28,767	4,241	0,000	23,956	0,000	5,901	0,000	30,494	0,000

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
Ir21a	RPRC001826	TCONS_00151191	186,836	116,423	96,090	8,702	4,964	4,964	14,701	4,201	175,840	68,151	76,838	21,067	144,476	65,036	90,066	3,755
Ir25a	RPRC000589	TCONS_00048737	319,321	97,006	256,565	60,245	139,001	0,000	16,334	0,000	287,275	103,138	267,693	41,860	298,927	104,523	234,597	8,451
Ir40a	RPRC000383	TCONS_00070191	11,381	11,381	28,084	17,203	19,857	9,929	3,267	0,000	40,150	25,641	4,883	4,883	8,215	7,882	51,118	30,022
Ir41a	RPRC000191	TCONS_00047870	105,439	0,000	662,140	366,889	714,862	0,000	2321,120	0,000	873,471	489,186	79,676	0,000	118,884	0,000	1191,950	733,779
Ir41b	RPRC000489	TCONS_00001645	82,540	65,787	178,075	51,798	109,156	44,078	214,186	121,452	87,979	42,057	204,481	70,865	72,825	45,810	293,041	77,764
Ir41c	RPRC013097	TCONS_00102111	34,410	25,396	168,533	91,258	0,000	0,000	0,000	0,000	218,892	117,455	28,758	21,153	32,546	17,033	306,383	173,845
Ir68a	RPRC000328	TCONS_00013041	405,374	223,569	146,745	118,760	4,964	0,000	17,968	0,000	171,594	91,635	294,315	215,758	345,528	226,652	7,807	7,785
Ir75a	RPRC002112	TCONS_00117873	197,135	84,210	310,033	40,456	2264,140	0,000	1485,320	8,285	414,955	57,569	129,846	52,512	141,824	59,027	533,554	47,067
Ir75b	RPRC017348	TCONS_00058220	0,000	0,000	55,281	55,281	113,553	113,553	35,777	35,777	73,708	73,708	0,000	0,000	0,000	0,000	110,561	110,561
Ir75d	RPRC000105	TCONS_00151562	127,949	81,664	139,292	50,150	37,232	19,860	11,434	1,633	173,483	64,798	97,540	56,511	163,755	59,567	79,024	62,830
Ir75e	RPRC000047	TCONS_00151556	103,898	51,276	2185,730	2162,470	39,715	4,347	13,068	0,000	142,041	109,317	2841,540	2808,160	99,694	58,151	4275,980	4259,920
Ir75f	RPRC002124	TCONS_00151555	68,059	62,055	86,758	38,967	2,482	0,000	6,534	0,000	136,622	75,582	24,428	17,744	83,627	38,920	74,320	62,149
Ir75k	RPRC000314	TCONS_00151227	7,289	6,457	10,762	10,505	2,482	0,000	1,633	0,000	11,368	11,368	7,841	6,944	14,407	13,734	0,000	0,000
Ir75m	RPRC000247	TCONS_00138763	147,096	18,611	133,991	92,924	44,679	0,000	24,502	3,267	208,309	101,021	68,409	35,286	184,989	88,200	45,099	28,060
Ir75n	RPRC000007	TCONS_00138764	80,643	65,762	28,463	22,705	0,000	0,000	11,434	2,796	36,527	29,148	55,186	44,967	61,158	51,301	15,252	8,572
Ir75o	RPRC017345	TCONS_00040607	391,321	391,321	421,620	421,620	96,804	96,804	44,103	44,103	592,156	592,156	230,884	230,884	390,010	390,010	454,540	454,540
Ir75p	RPRC017346	TCONS_00031463	367,486	63,808	235,694	31,620	2015,510	42,679	1992,800	79,633	287,750	47,203	271,499	37,495	316,903	51,328	205,069	24,391
Ir76b	RPRC000469	TCONS_00004096	0,556	0,556	4,383	4,383	19,857	19,857	8,167	8,167	6,105	6,105	0,109	0,109	3,443	3,443	2,434	2,434
IR8a	RPRC002763	TCONS_00036964	167,568	167,568	110,115	110,115	14,893	14,893	3,267	3,267	227,427	227,427	31,105	31,105	192,276	192,276	3,246	3,246
Ir93a	RPRC008486	TCONS_00049525	95,560	1,901	192,131	152,107	324,388	173,752	1326,490	40,694	258,296	202,758	61,585	1,319	69,198	1,326	341,425	303,464
Ir101	RPRC005659	TCONS_00125690	25,314	25,313	704,949	696,014	7,446	7,337	35,936	35,936	33,363	21,450	923,445	923,445	31,036	22,101	1373,140	1373,140
Ir102	RPRC000847	TCONS_00031476	112,738	13,140	161,433	12,355	29,786	0,000	49,003	26,109	244,755	19,119	45,648	6,113	81,733	18,923	272,138	0,003
Ir103	RPRC000379	TCONS_00081598	257,554	158,728	469,542	2,382	1180,020	0,000	782,742	0,000	540,789	54,598	256,969	54,397	342,234	81,746	512,170	0,000
Ir104	RPRC002996	TCONS_00035472	544,230	461,780	190,428	54,584	12,411	7,995	8,167	8,167	409,527	217,510	207,198	163,122	449,643	275,332	25,801	20,285
Ir105	RPRC000538	TCONS_00031472	112,738	76,677	161,433	88,417	29,786	0,000	49,003	0,000	244,755	150,921	45,648	18,086	81,733	43,573	272,138	166,364
Ir106	RPRC000846	TCONS_00031462	367,486	227,925	235,694	124,269	2015,510	1372,89	1992,800	1707,84	287,750	210,171	271,499	107,471	316,903	178,5	205,069	119,464
Ir107	RPRC017366	TCONS_00056479	267,532	8,099	776,565	363,201	183,009	85,493	173,145	100,066	310,027	40,255	903,748	449,413	324,221	16,804	1172,220	700,894

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
Ir21a	RPRC001826	TCONS_00151191	186,836	116,423	96,090	8,702	4,964	4,964	14,701	4,201	175,840	68,151	76,838	21,067	144,476	65,036	90,066	3,755
Ir25a	RPRC000589	TCONS_00048737	319,321	97,006	256,565	60,245	139,001	0,000	16,334	0,000	287,275	103,138	267,693	41,860	298,927	104,523	234,597	8,451
Ir40a	RPRC000383	TCONS_00070191	11,381	11,381	28,084	17,203	19,857	9,929	3,267	0,000	40,150	25,641	4,883	4,883	8,215	7,882	51,118	30,022
Ir41a	RPRC000191	TCONS_00047870	105,439	0,000	662,140	366,889	714,862	0,000	2321,120	0,000	873,471	489,186	79,676	0,000	118,884	0,000	1191,950	733,779
Ir41b	RPRC000489	TCONS_00001645	82,540	65,787	178,075	51,798	109,156	44,078	214,186	121,452	87,979	42,057	204,481	70,865	72,825	45,810	293,041	77,764
Ir41c	RPRC013097	TCONS_00102111	34,410	25,396	168,533	91,258	0,000	0,000	0,000	0,000	218,892	117,455	28,758	21,153	32,546	17,033	306,383	173,845
Ir68a	RPRC000328	TCONS_00013041	405,374	223,569	146,745	118,760	4,964	0,000	17,968	0,000	171,594	91,635	294,315	215,758	345,528	226,652	7,807	7,785
Ir75a	RPRC002112	TCONS_00117873	197,135	84,210	310,033	40,456	2264,140	0,000	1485,320	8,285	414,955	57,569	129,846	52,512	141,824	59,027	533,554	47,067
Ir75b	RPRC017348	TCONS_00058220	0,000	0,000	55,281	55,281	113,553	113,553	35,777	35,777	73,708	73,708	0,000	0,000	0,000	0,000	110,561	110,561
Ir75d	RPRC000105	TCONS_00151562	127,949	81,664	139,292	50,150	37,232	19,860	11,434	1,633	173,483	64,798	97,540	56,511	163,755	59,567	79,024	62,830
Ir75e	RPRC000047	TCONS_00151556	103,898	51,276	2185,730	2162,470	39,715	4,347	13,068	0,000	142,041	109,317	2841,540	2808,160	99,694	58,151	4275,980	4259,920
Ir75f	RPRC002124	TCONS_00151555	68,059	62,055	86,758	38,967	2,482	0,000	6,534	0,000	136,622	75,582	24,428	17,744	83,627	38,920	74,320	62,149
Ir75k	RPRC000314	TCONS_00151227	7,289	6,457	10,762	10,505	2,482	0,000	1,633	0,000	11,368	11,368	7,841	6,944	14,407	13,734	0,000	0,000
Ir75m	RPRC000247	TCONS_00138763	147,096	18,611	133,991	92,924	44,679	0,000	24,502	3,267	208,309	101,021	68,409	35,286	184,989	88,200	45,099	28,060
Ir75n	RPRC000007	TCONS_00138764	80,643	65,762	28,463	22,705	0,000	0,000	11,434	2,796	36,527	29,148	55,186	44,967	61,158	51,301	15,252	8,572
Ir75o	RPRC017345	TCONS_00040607	391,321	391,321	421,620	421,620	96,804	96,804	44,103	44,103	592,156	592,156	230,884	230,884	390,010	390,010	454,540	454,540
Ir75p	RPRC017346	TCONS_00031463	367,486	63,808	235,694	31,620	2015,510	42,679	1992,800	79,633	287,750	47,203	271,499	37,495	316,903	51,328	205,069	24,391
Ir76b	RPRC000469	TCONS_00004096	0,556	0,556	4,383	4,383	19,857	19,857	8,167	8,167	6,105	6,105	0,109	0,109	3,443	3,443	2,434	2,434
IR8a	RPRC002763	TCONS_00036964	167,568	167,568	110,115	110,115	14,893	14,893	3,267	3,267	227,427	227,427	31,105	31,105	192,276	192,276	3,246	3,246
Ir93a	RPRC008486	TCONS_00049525	95,560	1,901	192,131	152,107	324,388	173,752	1326,490	40,694	258,296	202,758	61,585	1,319	69,198	1,326	341,425	303,464
Ir101	RPRC005659	TCONS_00125690	25,314	25,313	704,949	696,014	7,446	7,337	35,936	35,936	33,363	21,450	923,445	923,445	31,036	22,101	1373,140	1373,140
Ir102	RPRC000847	TCONS_00031476	112,738	13,140	161,433	12,355	29,786	0,000	49,003	26,109	244,755	19,119	45,648	6,113	81,733	18,923	272,138	0,003
Ir103	RPRC000379	TCONS_00081598	257,554	158,728	469,542	2,382	1180,020	0,000	782,742	0,000	540,789	54,598	256,969	54,397	342,234	81,746	512,170	0,000
Ir104	RPRC002996	TCONS_00035472	544,230	461,780	190,428	54,584	12,411	7,995	8,167	8,167	409,527	217,510	207,198	163,122	449,643	275,332	25,801	20,285
Ir105	RPRC000538	TCONS_00031472	112,738	76,677	161,433	88,417	29,786	0,000	49,003	0,000	244,755	150,921	45,648	18,086	81,733	43,573	272,138	166,364
Ir106	RPRC000846	TCONS_00031462	367,486	227,925	235,694	124,269	2015,510	1372,89	1992,800	1707,84	287,750	210,171	271,499	107,471	316,903	178,5	205,069	119,464
Ir107	RPRC017366	TCONS_00056479	267,532	8,099	776,565	363,201	183,009	85,493	173,145	100,066	310,027	40,255	903,748	449,413	324,221	16,804	1172,220	700,894

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
Gr1	RPRC001549	TCONS_00114354	1730,410	1670,23	2997,100	2,617	933,292	329,906	645,209	273,175	3820,270	3491,63	1329,470	1275,24	1175,970	1000,94	5372,660	5148,41
Gr2	RPRC001795	TCONS_00146906	20,048	20,048	27,086	23,662	7,446	4,964	3,267	0,000	25,232	20,956	24,247	23,959	25,750	23,537	22,719	20,299
Gr3	RPRC002023	TCONS_00175448	118,004	0,000	373,805	0,125	5520,840	2566,98	2173,380	613,956	131,904	0,167	445,172	0,000	81,610	0,000	702,394	0,250
Gr4	RPRC000056	TCONS_00148495	23,117	23,117	11,165	10,913	327,645	327,645	248,283	248,283	15,156	14,821	15,141	15,141	19,883	19,631	5,680	5,680
Gr5	RPRC001197	TCONS_00073945	55,430	55,430	38,609	38,609	54,608	54,608	73,505	73,505	40,438	40,438	47,994	47,994	52,124	52,124	28,399	28,399
Gr6	RPRC000553	TCONS_00073943	17,028	11,047	14,666	7,970	230,841	131,693	380,592	277,443	13,779	1,233	17,128	16,758	17,095	13,493	12,171	0,000
Gr7	RPRC001196	TCONS_00073939	48,867	48,005	77,199	77,199	324,597	312,841	191,591	166,980	112,864	112,864	22,646	22,072	41,691	41,260	119,883	119,883
Gr8	RPRC001195	TCONS_00073937	29,950	29,950	41,878	38,915	170,055	153,668	209,287	206,012	76,684	42,734	29,119	29,119	56,021	53,059	1,663	1,663
Gr9	RPRC000388	TCONS_00073938	1,914	1,914	3,109	3,109	30,638	30,638	36,885	36,885	0,261	0,261	5,161	5,161	2,120	2,120	3,893	3,893
Gr10	RPRC000106	TCONS_00073936	79,049	46,659	6,087	6,087	27,304	17,008	65,338	51,289	42,192	22,111	18,623	17,111	45,206	29,010	0,811	0,811
Gr11	RPRC006740	TCONS_00073929	170,944	170,944	256,337	256,337	19,857	19,857	17,968	17,968	164,705	164,705	291,040	291,040	127,939	127,939	427,740	427,740
Gr12	RPRC002133	TCONS_00045622	11,407	11,407	14,180	13,506	12,411	4,726	1,633	1,633	18,578	18,578	7,934	7,035	7,713	7,039	24,342	24,342
Gr13	RPRC002134	TCONS_00045621	0,493	0,493	0,811	0,811	2,482	2,482	3,267	3,267	1,082	1,082	0,328	0,328	0,246	0,246	1,623	1,623
Gr15	RPRC001115	TCONS_00063607	27,613	22,291	35,847	17,745	0,000	0,000	4,900	4,900	49,574	35,767	16,630	2,753	28,397	18,754	42,511	20,272
Gr16	RPRC000730	TCONS_00014459	14,085	11,457	13,218	7,474	0,000	0,000	55,537	23,227	3,916	3,916	23,097	13,687	20,260	13,202	0,000	0,000
Gr17	RPRC001279	TCONS_00084114	22,813	22,813	66,814	44,846	7,446	7,446	9,801	9,801	78,762	52,146	25,533	22,857	18,100	16,093	120,242	80,318
Gr18	RPRC000336	TCONS_00093103	14,338	14,268	10,608	9,981	29,786	29,786	4,900	4,900	8,463	7,627	15,240	15,193	12,991	12,956	9,572	8,318
Gr19	RPRC000763	TCONS_00017128	88,094	72,923	143,441	116,560	0,000	0,000	39,203	0,000	69,713	48,600	180,270	155,428	61,618	50,860	251,739	204,322
Gr20	RPRC000451	TCONS_00096109	225,457	225,457	402,192	402,192	171,269	171,269	298,920	298,920	517,329	517,329	169,232	169,232	430,704	430,704	168,433	168,433
Gr21	RPRC000130	TCONS_00158462	51,032	42,328	5,192	0,000	0,000	0,000	0,000	0,000	32,056	27,671	8,888	0,547	29,085	21,164	3,246	0,000
Gr22	RPRC000073	TCONS_00066051	28,383	28,383	9,996	9,996	22,339	22,339	1,633	1,633	21,053	21,053	11,196	11,196	20,536	20,536	7,303	7,303
Gr24	RPRC001956	TCONS_00165992	5,419	5,419	3,109	3,109	2,482	2,482	3,267	3,267	0,000	0,000	7,759	7,759	3,873	3,873	3,893	3,893
Gr25	RPRC000216	TCONS_00024779	4,345	4,345	7,128	7,128	0,000	0,000	1,633	1,633	5,817	5,817	6,584	6,584	2,974	2,974	12,654	12,654
Gr26	RPRC000368	TCONS_00058943	18,897	3,607	70,416	41,948	772,059	0,031	1353,260	141,224	66,449	36,513	40,037	21,822	55,486	25,098	48,757	37,307
Gr27	RPRC000290	TCONS_00043967	1,604	1,604	93,380	93,380	17,376	17,376	66,971	66,971	122,774	122,774	2,803	2,803	2,494	2,494	183,377	183,377
Gr28	RPRC000068	TCONS_00043964	10,611	8,545	67,854	55,242	225,878	97,189	217,248	95,104	73,522	72,145	24,024	7,207	60,670	58,703	24,979	1,623

Nom	Gène	Transcrit le plus représentatif	RPKM Organes chimioS <i>R. prolixus</i>		RPKM Organes chimioS <i>R. robustus</i>		RPKM Tête <i>R. prolixus</i>		RPKM Tête <i>R. robustus</i>		RPKM Organes chimioS Femelles		RPKM Organes chimioS Mâles		RPKM Organes chimioS Elevages		RPKM Organes chimioS Sauvages	
			Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit	Gène	Transcrit
To1	XLOC_035880	TCONS_00043551	99,526	99,103	333,671	145,323	12,411	12,411	0,000	0,000	418,296	211,659	92,949	48,174	81,448	73,244	603,971	243,261
To2	XLOC_035881	TCONS_00043555	60,848	52,067	114,283	84,106	0,000	0,000	1,633	0,000	109,767	80,556	83,176	66,297	83,873	71,214	121,669	77,852
To3	XLOC_054441	TCONS_00070531	71,941	17,483	129,870	80,245	7679,800	102,344	2048,340	68,926	69,568	34,132	151,553	84,516	77,753	20,715	176,176	136,542
To4	XLOC_066090	TCONS_00086633	6,165	5,836	9,021	7,804	12,411	2,482	14,701	4,900	9,689	8,066	6,449	6,230	8,452	8,288	7,303	4,868
To5	XLOC_066906	TCONS_00087871	148,662	134,081	100,799	23,504	1603,480	203,010	1924,190	248,576	156,699	45,792	76,807	74,934	139,195	90,544	71,869	0,000
To6	XLOC_069601	TCONS_00091589	27,803	0,015	102,970	77,832	10797,400	1,190	8954,530	67,103	26,063	0,021	129,766	103,765	29,282	0,023	175,180	155,633
To7	XLOC_069603	TCONS_00091598	49,126	0,000	111,079	73,026	491,468	357,244	99,640	0,183	46,424	1,351	134,432	96,017	59,571	0,000	152,142	146,052
To8	XLOC_069604	TCONS_00091603	198,868	64,336	1875,800	1507,080	652,808	21,662	1109,110	2,338	531,161	109,680	2102,480	1942,650	269,048	122,945	3412,370	2832,610
To9	XLOC_069608	TCONS_00091622	70,978	48,402	90,675	77,816	414,521	0,011	449,196	258,291	127,825	122,329	40,394	13,694	95,901	77,323	60,527	49,388
To10	XLOC_069643	TCONS_00091774	763,003	762,985	327,586	11,965	270,554	0,000	168,244	3,148	481,739	480,647	463,711	43,964	394,918	393,458	628,340	0,000
To11	XLOC_069644	TCONS_00091778	207,479	202,816	102,807	93,842	417,003	15,562	315,254	15,935	170,311	167,941	105,084	92,391	170,362	163,976	72,369	62,547
To12	XLOC_069645	TCONS_00091780	24,267	20,674	5,648	4,602	255,662	21,813	633,775	37,712	4,679	4,138	19,029	15,780	17,375	14,939	0,811	0,000
To13	XLOC_069646	TCONS_00091788	261,885	249,625	1120,570	1114,18	121,626	34,7502	423,061	90,719	442,065	437,849	1226,620	1214,14	381,071	368,555	1740,880	1740,88
To14	XLOC_069647	TCONS_00091792	114,105	38,733	70,876	12,547	9782,190	505,910	12428,900	1964,250	89,583	5,109	80,988	37,442	80,873	27,388	94,112	9,049
To15	XLOC_084661	TCONS_00114299	59,774	51,222	48,368	13,366	9,929	9,920	119,241	5,447	64,962	32,054	39,377	19,915	67,547	35,731	21,415	6,491
To16	XLOC_084662	TCONS_00114300	1,642	1,642	0,423	0,423	9,929	9,929	3,267	3,267	0,000	0,000	1,659	1,659	1,244	1,244	0,000	0,000
To17	XLOC_084663	TCONS_00114302	158,594	114,700	61,076	17,986	81,911	0,841	11,434	0,000	108,772	46,628	78,391	53,820	103,454	73,401	73,838	3,871
To18	XLOC_084696	TCONS_00114408	273,262	198,919	284,520	204,663	255,662	15,616	207,447	18,022	381,387	285,409	180,148	120,088	231,068	180,765	380,167	246,714
To19	XLOC_084726	TCONS_00114510	147,832	38,529	156,868	103,356	563,986	8,456	363,423	0,540	268,965	160,066	38,748	3,428	95,373	28,301	270,823	188,639
To20	XLOC_084727	TCONS_00114514	27,803	15,331	11,818	6,886	556,004	57,217	2518,770	354,547	15,511	5,599	18,782	13,803	20,852	13,625	9,737	1,853
To11	XLOC_086924	TCONS_00117821	57,021	0,000	91,597	56,320	1752,400	476,516	1355,760	1073,500	113,062	75,094	47,081	0,000	51,955	0,000	136,305	112,640
To22	XLOC_104933	TCONS_00143435	0,164	0,164	0,545	0,545	0,000	0,000	0,000	0,000	0,444	0,444	0,391	0,391	0,627	0,627	0,000	0,000
To23	XLOC_128696	TCONS_00177776	79,643	18,253	86,279	73,761	186,162	186,162	147,010	71,901	130,824	81,564	37,310	28,952	58,918	24,653	134,364	116,469
To24	XLOC_128704	TCONS_00177812	351,053	149,126	582,376	303,651	17362,700	285,378	27917,100	197,883	760,641	367,263	249,896	137,023	389,295	257,457	737,216	241,514

Tableau 2 : Présence/Absence des transcrits de gènes chimiosensoriels (CSPs, OBP, ORs, IRs, GRs) et des protéines de la famille des takeout estimée à partir des données de RPKM de Tophat.

Nom	Gène	Transcrit maj de l'ensemble des ech	Transcrit maj de Rpro ds organes chimioS	Transcrit maj de Rrob ds organes chimioS	Transcrit maj de Rpro ds organes tête	Transcrit maj de Rrob ds organes tête	Transcrit maj de F ds organes chimioS	Transcrit maj de M ds organes chimioS	Transcrit maj de elev ds organes chimioS	Transcrit maj de sauv ds organes chimioS
CSP1	RPRC013223	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919	TCONS_00061919
CSP2	RPRC013221	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922	TCONS_00061922
CSP3	RPRC013280	TCONS_00061923	TCONS_00061923	TCONS_00061923	TCONS_00061923	TCONS_00061925	TCONS_00061924	TCONS_00061923	TCONS_00061923	TCONS_00061924
CSP4	RPRC013285	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926	TCONS_00061926
CSP5	RPRC013218	TCONS_00061929	TCONS_00061930	TCONS_00061929	TCONS_00061928	TCONS_00061928	TCONS_00061929	TCONS_00061928	TCONS_00061929	TCONS_00061929
CSP6	RPRC000332	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932	TCONS_00061932
CSP7	RPRC010107	TCONS_00095976	TCONS_00095976	TCONS_00095976	TCONS_00095977	TCONS_00095977	TCONS_00095976	TCONS_00095976	TCONS_00095976	TCONS_00095976
CSP8	RPRC000631	TCONS_00002593	TCONS_00002593	TCONS_00002593	TCONS_00002591	TCONS_00002591	TCONS_00002593	TCONS_00002593	TCONS_00002593	TCONS_00002593
CSP9	RPRC000470	TCONS_00093752	TCONS_00093752	TCONS_00093752	TCONS_00093751	TCONS_00093750	TCONS_00093752	TCONS_00093752	TCONS_00093752	TCONS_00093752
CSP10	RPRC007979	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934	TCONS_00171934
CSP11	RPRC011935	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364	TCONS_00177364
CSP12	RPRC011928	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373	TCONS_00177373
CSP13	RPRC011926	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177804	TCONS_00177806
CSP14	RPRC006791	TCONS_00099137	TCONS_00099137	TCONS_00099137	Non détecté	Non détecté	TCONS_00099137	TCONS_00099137	TCONS_00099137	TCONS_00099137
CSP15	RPRC000546	TCONS_00025500	TCONS_00025500	TCONS_00025500	Non détecté	Non détecté	TCONS_00025500	TCONS_00025500	TCONS_00025500	Non détecté
CSP16	RPRC000327	TCONS_00150597	TCONS_00150598	TCONS_00150597	TCONS_00150598	TCONS_00150598	TCONS_00150597	TCONS_00150598	TCONS_00150597	TCONS_00150597
CSP17	RPRC007134	TCONS_00000503	TCONS_00000503	TCONS_00000503	TCONS_00000503	TCONS_00000503	TCONS_00000503	TCONS_00000503	TCONS_00000503	Non détecté
CSP18	RPRC000242	TCONS_00077713	TCONS_00077713	TCONS_00077713	Non détecté	Non détecté	TCONS_00077713	TCONS_00077713	TCONS_00077713	TCONS_00077713
CSP19	RPRC000126	TCONS_00078853	Non détecté	TCONS_00078853	TCONS_00078853	TCONS_00078853	TCONS_00078853	TCONS_00078853	TCONS_00078853	Non détecté
OBP1	RPRC010284	TCONS_00154195	TCONS_00154197	TCONS_00154195	TCONS_00154195	TCONS_00154195	TCONS_00154195	TCONS_00154195	TCONS_00154197	TCONS_00154195
OBP2	RPRC000378	Non détecté	Non détecté	Non détecté	Non détecté	Non détecté	Non détecté	Non détecté	Non détecté	Non détecté
OBP3	RPRC000527	TCONS_00163866	TCONS_00163866	TCONS_00163866	TCONS_00163863	TCONS_00163863	TCONS_00163866	TCONS_00163866	TCONS_00163866	TCONS_00163866
OBP4	RPRC000274	TCONS_00054965	TCONS_00054965	TCONS_00054965	TCONS_00054964	TCONS_00054964	TCONS_00054964	TCONS_00054965	TCONS_00054965	TCONS_00054965
OBP5	XLOC_059757	TCONS_00077871	TCONS_00077871	TCONS_00077871	TCONS_00077871	TCONS_00077871	TCONS_00077871	TCONS_00077873	TCONS_00077871	TCONS_00077871
OBP6	RPRC004413	TCONS_00168603	TCONS_00168603	TCONS_00168603	TCONS_00168604	TCONS_00168604	TCONS_00168603	TCONS_00168603	TCONS_00168603	TCONS_00168603
OBP7	RPRC000311	TCONS_00168607	TCONS_00168607	TCONS_00168607	TCONS_00168607	TCONS_00168607	Non détecté	TCONS_00168607	TCONS_00168607	Non détecté
OBP8	RPRC000564	TCONS_00168610	TCONS_00168610	TCONS_00168610	TCONS_00168609	TCONS_00168609	TCONS_00168610	TCONS_00168610	TCONS_00168610	TCONS_00168610
OBP9	RPRC000194	TCONS_00168611	TCONS_00168611	Non détecté	Non détecté	Non détecté	TCONS_00168611	Non détecté	TCONS_00168611	Non détecté
OBP10	RPRC000560	TCONS_00168612	TCONS_00168612	TCONS_00168662	TCONS_00168612	TCONS_00168612	TCONS_00168612	TCONS_00168660	TCONS_00168612	TCONS_00168662
OBP11	RPRC004408	TCONS_00168667	TCONS_00168667	TCONS_00168667	TCONS_00168669	TCONS_00168669	TCONS_00168667	TCONS_00168667	TCONS_00168667	TCONS_00168667
OBP12	RPRC000492	TCONS_00168613	TCONS_00168613	TCONS_00168613	TCONS_00168614	TCONS_00168614	TCONS_00168613	TCONS_00168613	TCONS_00168613	TCONS_00168613
OBP13	RPRC000177	TCONS_00168615	TCONS_00168615	TCONS_00168615	TCONS_00168615	TCONS_00168616	TCONS_00168615	TCONS_00168615	TCONS_00168615	TCONS_00168615
OBP14	RPRC004407	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618	TCONS_00168618
OBP15	RPRC008413	TCONS_00067373	TCONS_00067373	TCONS_00067373	Non détecté	Non détecté	TCONS_00067373	TCONS_00067373	TCONS_00067373	Non détecté
OBP16	RPRC011610	TCONS_00059698	TCONS_00059699	TCONS_00059698	TCONS_00059698	TCONS_00059698	TCONS_00059699	TCONS_00059698	TCONS_00059698	TCONS_00059698
OBP17	RPRC000118	TCONS_00177251	TCONS_00177251	TCONS_00177251	TCONS_00177250	TCONS_00177250	TCONS_00177251	TCONS_00177251	TCONS_00177251	TCONS_00177251
OBP18	RPRC001924	TCONS_00160295	TCONS_00160295	TCONS_00160296	TCONS_00160296	TCONS_00160296	Non détecté	TCONS_00160296	TCONS_00160295	TCONS_00160296
OBP19	RPRC000593	TCONS_00160301	TCONS_00160301	TCONS_00160301	TCONS_00160300	TCONS_00160299	TCONS_00160301	TCONS_00160301	TCONS_00160301	Non détecté
OBP20	RPRC007008	TCONS_00160304	TCONS_00160304	TCONS_00160304	TCONS_00160306	TCONS_00160302	TCONS_00160304	TCONS_00160306	TCONS_00160304	TCONS_00160304
OBP21	RPRC001925	TCONS_00160312	TCONS_00160312	TCONS_00160312	TCONS_00160314	TCONS_00160314	TCONS_00160312	TCONS_00160312	TCONS_00160312	TCONS_00160312
OBP22	RPRC000115	TCONS_00069219	TCONS_00069219	TCONS_00069218	TCONS_00069222	TCONS_00069221	TCONS_00069219	TCONS_00069218	TCONS_00069219	Non détecté
OBP23	RPRC000174	TCONS_00069224	TCONS_00069224	TCONS_00069224	TCONS_00069226	TCONS_00069225	TCONS_00069224	TCONS_00069224	TCONS_00069224	TCONS_00069224
OBP24	RPRC009905	TCONS_00069231	TCONS_00069231	TCONS_00069231	TCONS_00069232	TCONS_00069232	TCONS_00069231	TCONS_00069231	TCONS_00069231	TCONS_00069231
OBP25	RPRC000257	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263	TCONS_00069263
OBP26	RPRC000182	TCONS_00069276	TCONS_00069276	TCONS_00069276	TCONS_00069280	TCONS_00069278	TCONS_00069281	TCONS_00069276	TCONS_00069276	TCONS_00069281
OBP27	RPRC009914	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282	TCONS_00069282
OBP28	RPRC009906	TCONS_00069213	TCONS_00069214	TCONS_00069214	TCONS_00069214	TCONS_00069214	TCONS_00069214	TCONS_00069214	TCONS_00069214	TCONS_00069214

Nom	Gène	Transcrit maj de l'ensemble des ech	Transcrit maj de Rpro ds organes chimioS	Transcrit maj de Rrob ds organes chimioS	Transcrit maj de Rpro ds organes tête	Transcrit maj de Rrob ds organes tête	Transcrit maj de F ds organes chimioS	Transcrit maj de M ds organes chimioS	Transcrit maj de elev ds organes chimioS	Transcrit maj de sauv ds organes chimioS
Or1	RPRC000579	TCONS_00150419	TCONS_00150421	TCONS_00150419	TCONS_00150420	TCONS_00150420	TCONS_00150419	TCONS_00150421	TCONS_00150419	TCONS_00150417
Or3	RPRC000059	TCONS_00037174	TCONS_00037174	TCONS_00037174	Non détecté	TCONS_00037174	TCONS_00037174	TCONS_00037174	TCONS_00037174	TCONS_00037174
Or5	RPRC000083	TCONS_00135190	TCONS_00135190	TCONS_00135191	Non détecté	Non détecté	TCONS_00135190	TCONS_00135190	TCONS_00135190	TCONS_00135189
Or10	RPRC000601	TCONS_00164670	TCONS_00164668	TCONS_00164670	Non détecté	TCONS_00164670	TCONS_00164670	TCONS_00164668	TCONS_00164668	TCONS_00164670
Or12	RPRC000376	TCONS_00068718	TCONS_00068718	TCONS_00068718	Non détecté	TCONS_00068717	TCONS_00068718	TCONS_00068718	TCONS_00068718	Non détecté
Or13	RPRC000362	TCONS_00056495	TCONS_00056495	TCONS_00056495	TCONS_00056495	TCONS_00056494	TCONS_00056495	TCONS_00056495	TCONS_00056495	TCONS_00056495
Or14	RPRC000166	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558	TCONS_00035558
Or14	RPRC000547	TCONS_00143424	TCONS_00143424	TCONS_00143424	TCONS_00143423	TCONS_00143423	TCONS_00143424	TCONS_00143424	TCONS_00143424	TCONS_00143424
Or16	RPRC000420	TCONS_00126179	TCONS_00126179	TCONS_00126179	Non détecté	TCONS_00126180	TCONS_00126179	TCONS_00126179	TCONS_00126179	TCONS_00126179
Or18	RPRC000616	TCONS_00126117	TCONS_00126117	TCONS_00126117	Non détecté	Non détecté	TCONS_00126117	TCONS_00126117	TCONS_00126117	TCONS_00126117
Or19	RPRC000449	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227	TCONS_00109227
Or20	RPRC000573	TCONS_00037649	TCONS_00037649	TCONS_00037649	TCONS_00037649	TCONS_00037648	Non détecté	TCONS_00037649	TCONS_00037649	TCONS_00037648
Or21	RPRC0006570	TCONS_00047083	Non détecté	TCONS_00047083	Non détecté	Non détecté	Non détecté	TCONS_00047083	TCONS_00047083	Non détecté
Or22	RPRC000375	TCONS_00047096	TCONS_00047096	TCONS_00047096	Non détecté	Non détecté	TCONS_00047096	TCONS_00047096	TCONS_00047096	TCONS_00047096
Or23	RPRC000149	TCONS_00047098	TCONS_00047097	TCONS_00047098	Non détecté	TCONS_00047098	TCONS_00047098	TCONS_00047097	TCONS_00047098	TCONS_00047098
Or25	RPRC000029	TCONS_00062915	TCONS_00062916	TCONS_00062914	TCONS_00062913	TCONS_00062915	TCONS_00062914	TCONS_00062916	TCONS_00062916	TCONS_00062914
Or26	RPRC000619	TCONS_00062892	TCONS_00062892	TCONS_00062892	TCONS_00062893	TCONS_00062893	TCONS_00062892	TCONS_00062892	TCONS_00062892	TCONS_00062893
Or29	RPRC000125	TCONS_00019062	TCONS_00019062	TCONS_00019062	TCONS_00019063	TCONS_00019062	TCONS_00019062	TCONS_00019062	TCONS_00019062	TCONS_00019063
Or30	RPRC000209	TCONS_00040217	TCONS_00040217	TCONS_00040217	Non détecté	TCONS_00040217	TCONS_00040217	TCONS_00040217	TCONS_00040217	TCONS_00040217
Or31	RPRC000423	TCONS_00069234	TCONS_00069234	TCONS_00069234	TCONS_00069234	Non détecté	TCONS_00069234	TCONS_00069234	TCONS_00069234	TCONS_00069234
Or32	RPRC000043	TCONS_00059354	TCONS_00059355	TCONS_00059354	TCONS_00059354	TCONS_00059355	TCONS_00059355	TCONS_00059354	TCONS_00059354	TCONS_00059355
Or33	RPRC000235	TCONS_00059353	TCONS_00059352	TCONS_00059353	TCONS_00059353	TCONS_00059351	TCONS_00059353	TCONS_00059353	TCONS_00059352	TCONS_00059353
Or36	RPRC000337	TCONS_00036060	TCONS_00036058	TCONS_00036058	TCONS_00036060	TCONS_00036059	TCONS_00036058	TCONS_00036060	TCONS_00036058	TCONS_00036060
Or37	RPRC000456	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062	TCONS_00036062
Or38	RPRC000530	TCONS_00171686	TCONS_00171686	TCONS_00171686	TCONS_00171687	TCONS_00171687	TCONS_00171686	TCONS_00171686	TCONS_00171686	TCONS_00171686
Or39	RPRC000041	TCONS_00121637	TCONS_00121637	TCONS_00121636	TCONS_00121637	TCONS_00121637	TCONS_00121636	TCONS_00121636	TCONS_00121637	TCONS_00121636
Or40	RPRC000229	TCONS_00034389	TCONS_00034389	TCONS_00034389	TCONS_00034389	TCONS_00034390	TCONS_00034389	TCONS_00034389	TCONS_00034389	TCONS_00034389
Or41	RPRC000146	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697	TCONS_00081697
Or42	RPRC000108	TCONS_00006024	TCONS_00006020	TCONS_00006024	Non détecté	TCONS_00006024	TCONS_00006024	TCONS_00006020	TCONS_00006024	TCONS_00006020
Or43	RPRC000140	TCONS_00086186	TCONS_00086186	TCONS_00086186	TCONS_00086186	TCONS_00086186	TCONS_00086186	TCONS_00086186	TCONS_00086186	Non détecté
Or47	RPRC000413	TCONS_00086206	TCONS_00086206	TCONS_00086206	Non détecté	TCONS_00086206	TCONS_00086207	TCONS_00086206	TCONS_00086206	TCONS_00086206
Or49	RPRC000433	TCONS_00125502	TCONS_00125502	TCONS_00125502	Non détecté	Non détecté	TCONS_00125502	TCONS_00125502	TCONS_00125502	TCONS_00125502
Or51	RPRC000507	TCONS_00079525	TCONS_00079525	TCONS_00079524	TCONS_00079525	TCONS_00079525	TCONS_00079525	TCONS_00079524	TCONS_00079525	TCONS_00079524
Or52	RPRC000201	TCONS_00072697	TCONS_00072697	TCONS_00072697	TCONS_00072697	Non détecté	TCONS_00072697	TCONS_00072697	TCONS_00072697	TCONS_00072697
Or53	RPRC002122	TCONS_00147337	TCONS_00147337	TCONS_00147334	Non détecté	TCONS_00147332	TCONS_00147337	TCONS_00147335	TCONS_00147334	TCONS_00147335
Or54	RPRC000055	TCONS_00147331	TCONS_00147331	TCONS_00147331	Non détecté	TCONS_00147330	TCONS_00147331	TCONS_00147331	TCONS_00147331	TCONS_00147331
Or57	RPRC000318	TCONS_00086863	TCONS_00086862	TCONS_00086863	Non détecté	TCONS_00086863	TCONS_00086863	TCONS_00086863	TCONS_00086862	TCONS_00086863
Or58	RPRC000526	TCONS_00035788	TCONS_00035788	TCONS_00035788	TCONS_00035787	TCONS_00035787	TCONS_00035788	TCONS_00035788	TCONS_00035788	TCONS_00035784
Or59	RPRC004587	TCONS_00097321	TCONS_00097321	TCONS_00097321	TCONS_00097323	TCONS_00097321	TCONS_00097321	TCONS_00097321	TCONS_00097321	TCONS_00097321
Or60	RPRC002105	TCONS_00088042	Non détecté	TCONS_00088043	Non détecté	Non détecté	TCONS_00088043	TCONS_00088042	TCONS_00088043	TCONS_00088043
Or61	RPRC000208	TCONS_00088071	TCONS_00088070	TCONS_00088071	Non détecté	TCONS_00088070	TCONS_00088071	TCONS_00088071	TCONS_00088071	TCONS_00088071

Or64	RPRC000095	TCONS_00088063	TCONS_00088063	TCONS_00088063	TCONS_00088063	Non détecté	TCONS_00088063	TCONS_00088063	TCONS_00088063	Non détecté
Or65	RPRC000132	TCONS_00088060	TCONS_00088060	TCONS_00088060	TCONS_00088061	TCONS_00088060	TCONS_00088060	TCONS_00088060	TCONS_00088060	TCONS_00088060
Or66	RPRC000452	TCONS_00088055	TCONS_00088055	TCONS_00088055	TCONS_00088058	TCONS_00088057	TCONS_00088056	TCONS_00088055	TCONS_00088055	TCONS_00088056
Or67	RPRC000222	TCONS_00088052	TCONS_00088052	TCONS_00088039	TCONS_00088052	TCONS_00088052	TCONS_00088052	TCONS_00088052	TCONS_00088052	TCONS_00088039
Or68	RPRC000541	TCONS_00088049	TCONS_00088049	TCONS_00088049	TCONS_00088050	Non détecté	TCONS_00088048	TCONS_00088049	TCONS_00088049	TCONS_00088048
Or70	RPRC000459	TCONS_00088088	TCONS_00088088	TCONS_00088088	Non détecté	TCONS_00088088	TCONS_00088088	TCONS_00088088	TCONS_00088088	TCONS_00088088
Or71	RPRC000490	TCONS_00088086	TCONS_00088086	TCONS_00088086	TCONS_00088087	Non détecté	TCONS_00088086	TCONS_00088086	TCONS_00088086	TCONS_00088086
Or72	RPRC000053	TCONS_00088084	TCONS_00088084	Non détecté	TCONS_00088084	TCONS_00088084	TCONS_00088084	Non détecté	TCONS_00088084	Non détecté
Or73	RPRC000585	TCONS_00088082	TCONS_00088082	TCONS_00088082	TCONS_00088081	TCONS_00088081	TCONS_00088081	TCONS_00088082	TCONS_00088081	TCONS_00088082
Or74	RPRC000200	TCONS_00088092	TCONS_00088095	TCONS_00088092	Non détecté	TCONS_00088095	TCONS_00088095	TCONS_00088092	TCONS_00088095	TCONS_00088092
Or75	RPRC000245	TCONS_00088047	Non détecté	TCONS_00088046	TCONS_00088047	TCONS_00088047	TCONS_00088047	TCONS_00088046	TCONS_00088047	TCONS_00088047
Or77	RPRC000468	TCONS_00161181	TCONS_00161181	TCONS_00161181	TCONS_00161183	TCONS_00161183	TCONS_00161181	TCONS_00161181	TCONS_00161181	TCONS_00161181
Or79	RPRC000202	TCONS_00118300	TCONS_00118300	TCONS_00118300	TCONS_00118303	TCONS_00118300	TCONS_00118300	TCONS_00118300	TCONS_00118300	TCONS_00118301
Or80	RPRC000441	TCONS_00118296	TCONS_00118298	TCONS_00118296	TCONS_00118296	Non détecté	TCONS_00118296	TCONS_00118298	TCONS_00118298	TCONS_00118296
Or81	RPRC000363	TCONS_00118293	TCONS_00118293	TCONS_00118293	TCONS_00118295	Non détecté	TCONS_00118293	TCONS_00118293	TCONS_00118293	TCONS_00118293
Or82	RPRC002113	TCONS_00118290	TCONS_00118292	TCONS_00118290	TCONS_00118292	TCONS_00118290	TCONS_00118290	TCONS_00118292	TCONS_00118292	TCONS_00118290
Or83	RPRC0006981	TCONS_00036175	TCONS_00036175	TCONS_00036175	TCONS_00036174	TCONS_00036174	TCONS_00036175	TCONS_00036175	TCONS_00036175	TCONS_00036175
Or84	RPRC000431	TCONS_00036166	TCONS_00036165	TCONS_00036166	TCONS_00036167	TCONS_00036165	TCONS_00036165	TCONS_00036166	TCONS_00036166	TCONS_00036168
Or86	RPRC000079	TCONS_00036159	TCONS_00036159	TCONS_00036159	Non détecté	TCONS_00036160	TCONS_00036159	TCONS_00036159	TCONS_00036150	TCONS_00036159
Or87	RPRC000317	TCONS_00036157	TCONS_00036157	TCONS_00036157	TCONS_00036154	TCONS_00036157	TCONS_00036155	TCONS_00036157	TCONS_00036157	TCONS_00036157
Or88	RPRC000607	TCONS_00032721	TCONS_00032721	TCONS_00032721	TCONS_00032721	Non détecté	TCONS_00032721	TCONS_00032721	TCONS_00032721	TCONS_00032721
Or89	RPRC000165	TCONS_00174276	TCONS_00174276	Non détecté	Non détecté	Non détecté	TCONS_00174276	Non détecté	TCONS_00174276	Non détecté
Or90	RPRC000249	TCONS_00174272	TCONS_00174272	TCONS_00174274	TCONS_00174272	TCONS_00174274	TCONS_00174272	TCONS_00174272	TCONS_00174272	TCONS_00174273
Or91	RPRC000093	TCONS_00174269	TCONS_00174269	TCONS_00174269	Non détecté	Non détecté	TCONS_00174269	TCONS_00174269	TCONS_00174269	TCONS_00174269
Or97	RPRC0002120	TCONS_00143678	TCONS_00143678	TCONS_00143678	Non détecté	TCONS_00143679	TCONS_00143678	TCONS_00143678	TCONS_00143677	TCONS_00143678
Or99	RPRC000065	TCONS_00143672	TCONS_00143672	TCONS_00143672	TCONS_00143672	TCONS_00143669	TCONS_00143669	TCONS_00143672	TCONS_00143672	TCONS_00143672
Or100	RPRC000522	TCONS_00143666	TCONS_00143665	TCONS_00143666	TCONS_00143666	TCONS_00143666	TCONS_00143666	TCONS_00143666	TCONS_00143666	TCONS_00143666
Or101	RPRC000188	TCONS_00028771	TCONS_00028773	TCONS_00028771	TCONS_00028772	TCONS_00028775	TCONS_00028771	TCONS_00028773	TCONS_00028773	TCONS_00028771
Or104	RPRC000906	TCONS_00075071	TCONS_00075071	TCONS_00075071	Non détecté	TCONS_00075071	TCONS_00075071	TCONS_00075071	TCONS_00075071	TCONS_00075070
Or105	RPRC000255	TCONS_00121651	TCONS_00121646	TCONS_00121651	TCONS_00121648	TCONS_00121649	TCONS_00121658	TCONS_00121651	TCONS_00121652	TCONS_00121651
Or106	RPRC000371	TCONS_00139335	TCONS_00139334	TCONS_00139335	TCONS_00139335	TCONS_00139334	TCONS_00139334	TCONS_00139335	TCONS_00139334	TCONS_00139335
Or108	RPRC000301	TCONS_00139332	TCONS_00139332	TCONS_00139332	TCONS_00139333	TCONS_00139333	TCONS_00139332	TCONS_00139333	TCONS_00139332	TCONS_00139332
Or112	RPRC000555	TCONS_00120491	TCONS_00120494	TCONS_00120491	TCONS_00120491	TCONS_00120491	TCONS_00120491	TCONS_00120491	TCONS_00120494	TCONS_00120491
Or114	RPRC0004306	TCONS_00087779	TCONS_00087779	TCONS_00087779	TCONS_00087779	Non détecté	TCONS_00087779	TCONS_00087779	TCONS_00087779	TCONS_00087779
OrCo	RPRC000476	TCONS_00139686	TCONS_00139684	TCONS_00139684	TCONS_00139686	TCONS_00139685	TCONS_00139684	TCONS_00139684	TCONS_00139684	TCONS_00139684

Nom	Gène	Transcrit maj de l'ensemble des ech	Transcrit maj de Rpro ds organes chimioS	Transcrit maj de Rrob ds organes chimioS	Transcrit maj de Rpro ds organes tête	Transcrit maj de Rrob ds organes tête	Transcrit maj de F ds organes chimioS	Transcrit maj de M ds organes chimioS	Transcrit maj de elev ds organes chimioS	Transcrit maj de sauv ds organes chimioS
Ir21a	RPRC001826	TCONS_00151191	TCONS_00151191	TCONS_00151192	TCONS_00151191	TCONS_00151194	TCONS_00151192	TCONS_00151195	TCONS_00151191	TCONS_00151192
Ir25a	RPRC000589	TCONS_00048737	TCONS_00048736	TCONS_00048738	TCONS_00048732	TCONS_00048732	TCONS_00048737	TCONS_00048736	TCONS_00048737	TCONS_00048738
Ir40a	RPRC000383	TCONS_00070191	TCONS_00070191	TCONS_00070191	TCONS_00070191	TCONS_00070190	TCONS_00070191	TCONS_00070191	TCONS_00070191	TCONS_00070191
Ir41a	RPRC000191	TCONS_00047870	TCONS_00047869	TCONS_00047870	TCONS_00047868	TCONS_00047868	TCONS_00047870	TCONS_00047869	TCONS_00047869	TCONS_00047870
Ir41b	RPRC000489	TCONS_00001645	TCONS_00001645	TCONS_00001644	TCONS_00001644	TCONS_00001645	TCONS_00001645	TCONS_00001644	TCONS_00001645	TCONS_00001644
Ir41c	RPRC013097	TCONS_00102111	TCONS_00102111	TCONS_00102111	Non détecté	Non détecté	TCONS_00102111	TCONS_00102111	TCONS_00102111	TCONS_00102111
Ir68a	RPRC000328	TCONS_00013041	TCONS_00013041	TCONS_00013041	TCONS_00013045	TCONS_00013042	TCONS_00013041	TCONS_00013041	TCONS_00013041	TCONS_00013041
Ir75a	RPRC002112	TCONS_00117873	TCONS_00117873	TCONS_00117882	TCONS_00117875	TCONS_00117875	TCONS_00117882	TCONS_00117873	TCONS_00117873	TCONS_00117882
Ir75b	RPRC017348	TCONS_00058220	Non détecté	TCONS_00058220	TCONS_00058220	TCONS_00058220	TCONS_00058220	Non détecté	Non détecté	TCONS_00058220
Ir75d	RPRC000105	TCONS_00151562	TCONS_00151562	TCONS_00151560	TCONS_00151562	TCONS_00151561	TCONS_00151560	TCONS_00151562	TCONS_00151560	TCONS_00151562
Ir75e	RPRC000047	TCONS_00151556	TCONS_00151558	TCONS_00151556	TCONS_00151558	TCONS_00151558	TCONS_00151556	TCONS_00151556	TCONS_00151556	TCONS_00151556
Ir75f	RPRC002124	TCONS_00151555	TCONS_00151555	TCONS_00151554	TCONS_00151554	TCONS_00151553	TCONS_00151555	TCONS_00151555	TCONS_00151554	TCONS_00151555
Ir75k	RPRC000314	TCONS_00151227	TCONS_00151227	TCONS_00151227	TCONS_00151225	TCONS_00151225	TCONS_00151227	TCONS_00151227	TCONS_00151227	Non détecté
Ir75m	RPRC000247	TCONS_00138763	TCONS_00138760	TCONS_00138763	TCONS_00138762	TCONS_00138760	TCONS_00138763	TCONS_00138763	TCONS_00138763	TCONS_00138763
Ir75n	RPRC000007	TCONS_00138764	TCONS_00138764	TCONS_00138764	Non détecté	TCONS_00138765	TCONS_00138764	TCONS_00138764	TCONS_00138764	TCONS_00138764
Ir75o	RPRC017345	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607	TCONS_00040607
Ir76b	RPRC000469	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096	TCONS_00004096
IR8a	RPRC002763	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964	TCONS_00036964
Ir93a	RPRC008486	TCONS_00049525	TCONS_00049529	TCONS_00049525	TCONS_00049525	TCONS_00049529	TCONS_00049525	TCONS_00049524	TCONS_00049524	TCONS_00049525
Ir101	RPRC005659	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690	TCONS_00125690
Ir102	RPRC000847	TCONS_00031476	TCONS_00031472	TCONS_00031472	TCONS_00031477	TCONS_00031476	TCONS_00031476	TCONS_00031476	TCONS_00031472	TCONS_00031472
Ir103	RPRC000379	TCONS_00081598	TCONS_00081598	TCONS_00081599	TCONS_00081595	TCONS_00081595	TCONS_00081595	TCONS_00081591	TCONS_00081594	TCONS_00081599
Ir104	RPRC002996	TCONS_00035472	TCONS_00035472	TCONS_00035475	TCONS_00035472	TCONS_00035472	TCONS_00035472	TCONS_00035472	TCONS_00035472	TCONS_00035472
Ir105	RPRC000538	TCONS_00031472	TCONS_00031472	TCONS_00031472	TCONS_00031474	TCONS_00031474	TCONS_00031472	TCONS_00031472	TCONS_00031472	TCONS_00031472
Ir106	RPRC000846	TCONS_00031462	TCONS_00031463	TCONS_00031464	TCONS_00031462	TCONS_00031462	TCONS_00031462	TCONS_00031464	TCONS_00031462	TCONS_00031462
Ir107	RPRC017366	TCONS_00056479	TCONS_00056481	TCONS_00056479	TCONS_00056479	TCONS_00056479	TCONS_00056481	TCONS_00056479	TCONS_00056481	TCONS_00056479

Nom	Gène	Transcrit maj de l'ensemble des ech	Transcrit maj de Rpro ds organes chimioS	Transcrit maj de Rrob ds organes chimioS	Transcrit maj de Rpro ds organes tête	Transcrit maj de Rrob ds organes tête	Transcrit maj de F ds organes chimioS	Transcrit maj de M ds organes chimioS	Transcrit maj de elev ds organes chimioS	Transcrit maj de sauv ds organes chimioS
Gr1	RPRC001549	TCONS_00114354	TCONS_00114354	TCONS_00114354	TCONS_00114348	TCONS_00114348	TCONS_00114354	TCONS_00114354	TCONS_00114354	TCONS_00114354
Gr2	RPRC001795	TCONS_00146906	TCONS_00146906	TCONS_00146906	TCONS_00146906	TCONS_00146905	TCONS_00146906	TCONS_00146906	TCONS_00146906	TCONS_00146906
Gr3	RPRC002023	TCONS_00175456	TCONS_00175449	TCONS_00175453	TCONS_00175456	TCONS_00175456	TCONS_00175448	TCONS_00175451	TCONS_00175449	TCONS_00175453
Gr4	RPRC000056	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495	TCONS_00148495
Gr5	RPRC001197	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945	TCONS_00073945
Gr6	RPRC000553	TCONS_00073943	TCONS_00073943	TCONS_00073943	TCONS_00073943	TCONS_00073943	TCONS_00073942	TCONS_00073943	TCONS_00073943	TCONS_00073942
Gr7	RPRC001196	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939	TCONS_00073939
Gr8	RPRC001195	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937	TCONS_00073937
Gr9	RPRC000388	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938	TCONS_00073938
Gr10	RPRC000106	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936	TCONS_00073936
Gr11	RPRC006740	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929	TCONS_00073929
Gr12	RPRC002133	TCONS_00045622	TCONS_00045622	TCONS_00045622	TCONS_00045623	TCONS_00045622	TCONS_00045622	TCONS_00045622	TCONS_00045622	TCONS_00045622
Gr13	RPRC002134	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621	TCONS_00045621
Gr15	RPRC001115	TCONS_00063607	TCONS_00063607	TCONS_00063606	Non détecté	TCONS_00063607	TCONS_00063607	TCONS_00063607	TCONS_00063607	TCONS_00063606
Gr16	RPRC000730	TCONS_00014459	TCONS_00014459	TCONS_00014459	Non détecté	TCONS_00014458	TCONS_00014459	TCONS_00014459	TCONS_00014459	Non détecté
Gr17	RPRC001279	TCONS_00084114	TCONS_00084114	TCONS_00084114	TCONS_00084137	TCONS_00084137	TCONS_00084114	TCONS_00084114	TCONS_00084114	TCONS_00084114
Gr18	RPRC000336	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103	TCONS_00093103
Gr19	RPRC000763	TCONS_00017128	TCONS_00017128	TCONS_00017128	Non détecté	TCONS_00017129	TCONS_00017128	TCONS_00017128	TCONS_00017128	TCONS_00017128
Gr20	RPRC000451	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109	TCONS_00096109
Gr21	RPRC000130	TCONS_00158462	TCONS_00158462	TCONS_00158461	Non détecté	Non détecté	TCONS_00158462	TCONS_00158461	TCONS_00158462	TCONS_00158461
Gr22	RPRC000073	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051	TCONS_00066051
Gr24	RPRC001956	TCONS_00165992	TCONS_00165992	TCONS_00165992	TCONS_00165992	TCONS_00165992	Non détecté	TCONS_00165992	TCONS_00165992	TCONS_00165992
Gr25	RPRC000216	TCONS_00024779	TCONS_00024779	TCONS_00024779	Non détecté	TCONS_00024779	TCONS_00024779	TCONS_00024779	TCONS_00024779	TCONS_00024779
Gr26	RPRC000368	TCONS_00058943	TCONS_00058941	TCONS_00058943	TCONS_00058942	TCONS_00058942	TCONS_00058943	TCONS_00058943	TCONS_00058941	TCONS_00058943
Gr27	RPRC000290	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967	TCONS_00043967
Gr28	RPRC000068	TCONS_00043964	TCONS_00043964	TCONS_00043964	TCONS_00043964	TCONS_00043964	TCONS_00043964	TCONS_00043965	TCONS_00043964	TCONS_00043965

Nom	Gène	Transcrit maj de l'ensemble des ech	Transcrit maj de Rpro ds organes chimioS	Transcrit maj de Rrob ds organes chimioS	Transcrit maj de Rpro ds organes tête	Transcrit maj de Rrob ds organes tête	Transcrit maj de F ds organes chimioS	Transcrit maj de M ds organes chimioS	Transcrit maj de elev ds organes chimioS	Transcrit maj de sauv ds organes chimioS
To1	RPRC002524	TCONS_00043551	TCONS_00043551	TCONS_00043550	TCONS_00043551	Non détecté	TCONS_00043551	TCONS_00043551	TCONS_00043551	TCONS_00043550
To2	RPRC002529	TCONS_00043555	TCONS_00043555	TCONS_00043555	Non détecté	TCONS_00043556	TCONS_00043555	TCONS_00043555	TCONS_00043555	TCONS_00043555
To3	RPRC008276	TCONS_00070532	TCONS_00070536	TCONS_00070532	TCONS_00070533	TCONS_00070533	TCONS_00070532	TCONS_00070532	TCONS_00070532	TCONS_00070532
To4	RPRC005347	TCONS_00086631	TCONS_00086631	TCONS_00086631	TCONS_00086633	TCONS_00086633	TCONS_00086631	TCONS_00086631	TCONS_00086631	TCONS_00086631
To5	RPRC009613	TCONS_00087871	TCONS_00087871	TCONS_00087872	TCONS_00087872	TCONS_00087872	TCONS_00087872	TCONS_00087871	TCONS_00087871	TCONS_00087872
To6	RPRC010096	TCONS_00091589	TCONS_00091590	TCONS_00091589	TCONS_00091588	TCONS_00091587	TCONS_00091590	TCONS_00091589	TCONS_00091590	TCONS_00091589
To7	RPRC010085	TCONS_00091598	TCONS_00091599	TCONS_00091598	TCONS_00091598	TCONS_00091600	TCONS_00091597	TCONS_00091598	TCONS_00091599	TCONS_00091598
To	RPRC005775	TCONS_00091603	TCONS_00091606	TCONS_00091603	TCONS_00091604	TCONS_00091602	TCONS_00091607	TCONS_00091603	TCONS_00091603	TCONS_00091603
To9	RPRC005776	TCONS_00091622	TCONS_00091622	TCONS_00091622	TCONS_00091624	TCONS_00091622	TCONS_00091622	TCONS_00091625	TCONS_00091622	TCONS_00091622
To10	RPRC010202	TCONS_00091774	TCONS_00091774	TCONS_00091775	TCONS_00091775	TCONS_00091775	TCONS_00091774	TCONS_00091775	TCONS_00091774	TCONS_00091775
To11	RPRC010201	TCONS_00091778	TCONS_00091778	TCONS_00091778	TCONS_00091777	TCONS_00091777	TCONS_00091778	TCONS_00091778	TCONS_00091778	TCONS_00091778
To12	RPRC005774	TCONS_00091780	TCONS_00091780	TCONS_00091780	TCONS_00091781	TCONS_00091781	TCONS_00091780	TCONS_00091780	TCONS_00091780	TCONS_00091781
To13	RPRC005773	TCONS_00091788	TCONS_00091788	TCONS_00091788	TCONS_00091785	TCONS_00091784	TCONS_00091788	TCONS_00091788	TCONS_00091788	TCONS_00091788
To14	RPRC005893	TCONS_00091792	TCONS_00091791	TCONS_00091791	TCONS_00091790	TCONS_00091794	TCONS_00091791	TCONS_00091792	TCONS_00091791	TCONS_00091792
To15	RPRC002517	TCONS_00114299	TCONS_00114299	TCONS_00114295	TCONS_00114299	TCONS_00114297	TCONS_00114295	TCONS_00114299	TCONS_00114299	TCONS_00114295
To16	RPRC002506	TCONS_00114300	TCONS_00114300	TCONS_00114300	TCONS_00114300	TCONS_00114300	Non détecté	TCONS_00114300	TCONS_00114300	Non détecté
To17	RPRC002554	TCONS_00114302	TCONS_00114302	TCONS_00114303	TCONS_00114306	TCONS_00114306	TCONS_00114302	TCONS_00114302	TCONS_00114302	TCONS_00114303
To18	RPRC002531	TCONS_00114408	TCONS_00114408	TCONS_00114408	TCONS_00114410	TCONS_00114409	TCONS_00114408	TCONS_00114408	TCONS_00114408	TCONS_00114408
To19	RPRC008451	TCONS_00114510	TCONS_00114511	TCONS_00114510	TCONS_00114508	TCONS_00114508	TCONS_00114510	TCONS_00114511	TCONS_00114511	TCONS_00114510
To20	RPRC008440	TCONS_00114514	TCONS_00114514	TCONS_00114514	TCONS_00114515	TCONS_00114515	TCONS_00114516	TCONS_00114514	TCONS_00114514	TCONS_00114516
To11	RPRC006285	TCONS_00117821	TCONS_00117820	TCONS_00117821	TCONS_00117818	TCONS_00117821	TCONS_00117821	TCONS_00117819	TCONS_00117820	TCONS_00117821
To22	RPRC012883	TCONS_00143435	TCONS_00143435	TCONS_00143435	Non détecté	Non détecté	TCONS_00143435	TCONS_00143435	TCONS_00143435	Non détecté
To23	RPRC011983	TCONS_00177776	TCONS_00177777	TCONS_00177776	TCONS_00177776	TCONS_00177776	TCONS_00177776	TCONS_00177776	TCONS_00177777	TCONS_00177776
To24	RPRC012014	TCONS_00177812	TCONS_00177812	TCONS_00177812	TCONS_00177813	TCONS_00177813	TCONS_00177812	TCONS_00177812	TCONS_00177812	TCONS_00177812

Tableaux 3 : Identification de l'isoforme majoritaire (RPKM le plus élevé) pour les gènes chimiosensoriels et *takeout* et les différents échantillons regroupés en fonction de l'espèce, du tissu extrait, du sexe ou de l'environnement. La première colonne indique le transcrit identifié comme majoritaire sur l'ensemble des échantillons (somme des RPKM). Si celui-ci diffère pour un sous échantillon, sa case est colorée.

Chapitre 3 : Evolution du nombre de gènes chimiosensoriels : lien avec la capacité de domiciliation.

Résumé

Dans les chapitres précédents, nous avons pu mettre en évidence une différence d'expression des gènes chimiosensoriels entre des Triatomes (*T. brasiliensis* ou *Rhodnius*) provenant de milieux sylvatiques ou anthropisés (environnement domiciliés ou élevage). Ces premiers résultats suggèrent l'implication des gènes chimiosensoriels dans la capacité d'une espèce à s'adapter à un environnement anthropisé. Dans cette étude, les OBPs et les CSPs ont été annotées dans le transcriptome et dans le génome de 10 espèces différentes de *Rhodnius* afin de comparer leur évolution. Puis, nous avons réalisé un test de contrastes indépendants de la phylogénie afin de corrélérer le nombre de gènes et transcrits chimiosensoriels et la capacité de domiciliation des différentes espèces. Les résultats suggèrent que le mode d'évolution des protéines de liaison impliquant des phénomènes de duplication et de régulation d'expression est en lien avec l'adaptation des punaises hématophages à un nouvel environnement.

Discussion et Perspectives

Discussion

Dans un contexte où l'anthropisation des habitats naturels conduit les Triatominae à coloniser les milieux anthropiques, l'objectif de ce travail était d'aborder la réponse adaptative des vecteurs de la maladie de Chagas en réponse aux pressions anthropiques pour comprendre pourquoi certaines espèces présentent des populations domiciliées tandis que d'autres ne semblent pas capables de coloniser les milieux anthropisés. Nous nous sommes alors interrogés sur les gènes potentiellement impliqués dans la domiciliation des vecteurs et sur la possibilité que ces gènes soient les mêmes quelque soit l'espèce de Triatominae domiciliée. Le système chimiosensoriel ayant souvent été proposé comme acteur clé de la capacité d'un insecte à s'adapter à un nouvel habitat, nous avons exploré l'existence d'une relation entre l'expansion de ces gènes, leur expression et la capacité des Triatominae à s'adapter à un habitat anthropisé. Ainsi, les gènes chimiosensoriels de punaises hématophages d'habitats différents ont été répertoriés et comparés, en utilisant des données transcriptomiques et génomiques obtenues en séquençage à haut débit.

Dans un premier temps, nous évoquerons les limites de l'étude et les difficultés rencontrées. Puis, nous discuterons des conclusions et des hypothèses qui peuvent découler de ce travail. Nous verrons ce que notre modèle d'étude a pu apporter d'un point de vue méthodologique, dans le choix des stratégies d'assemblage transcriptomique et de l'analyse d'expression à partir de données de séquençage à haut débit. Une comparaison des résultats entre les analyses d'expression différentielle des deux modèles de Triatominae utilisés sera abordée afin d'évaluer la généralisation du processus de domiciliation. Enfin, nous verrons comment adaptation génétique et plasticité pourraient être impliquées dans la domiciliation en jouant notamment sur les répertoires de gènes chimiosensoriels.

Limites de l'étude

La principale limite de cette étude réside dans la difficulté de travailler avec des échantillons prélevés sur le terrain. En effet, l'échantillonnage n'est pas aisé et nécessite des campagnes de terrain dédiées. De plus, au Brésil, le travail sur le terrain et l'exportation de matériel biologique sont soumis à une réglementation administrativement lourde. Par ailleurs, le conditionnement des individus pour la préservation de l'ARN demande une procédure particulière qui n'est pas toujours compatible avec les objectifs prioritaires de recherche de nos collaborateurs à savoir la prévalence du parasite *T. cruzi* dans les différents échantillons de vecteurs. Ainsi, une campagne d'échantillonnage pour collecter davantage de réplicats avait été effectuée, mais la qualité de l'ARN à l'arrivée des punaises ne nous a pas permis d'inclure ces échantillons dans nos analyses de RNAseq, ni de les utiliser pour valider nos résultats d'expression différentielle par qPCR. Par ailleurs, les premières extractions ont été réalisées en 2011, il était alors nécessaire d'obtenir une quantité suffisante d'ARN pour le séquençage. A cette époque, la création de réplicats techniques était fortement recommandée. Le

regroupement de plusieurs individus par échantillon était également conseillé pour obtenir une mesure de l'expression « moyenne » d'une population. Le coût du séquençage était beaucoup plus élevé qu'aujourd'hui et les technologies Illumina moins développées, notamment le paired-end. Ces contraintes aussi bien biologiques que financières ont limité le nombre de réplicats biologiques utilisés dans les expériences de RNAseq. Le faible nombre de réplicats biologiques est courant dans les études de type RNAseq mais induit des difficultés pour les méthodes d'analyse différentielle (Zhang et al., 2014b). Ainsi les résultats obtenus dans de telles conditions doivent être interprétés avec prudence (Soneson and Delorenzi, 2013).

Par ailleurs, beaucoup de facteurs non contrôlés entrent en jeu dans notre plan d'échantillonnage pouvant influencer l'expression des gènes dans les échantillons. Par exemple, le comportement de recherche alimentaire des punaises hématophages dépend d'un facteur non contrôlable sur le terrain, celui de la durée qui sépare le moment de la collecte et celui du dernier repas sanguin de la punaise. De plus, le comportement de recherche alimentaire des Triatominae étant accentué en début de nuit, l'heure de collecte va influencer l'expression des gènes impliqués dans ce comportement. Par ailleurs, les pièges lumineux capturent des individus en déplacement, donc ayant un comportement qui diffère de ceux capturés dans les anfractuosités des murs des habitations humaines. Ce comportement engendre très probablement des modifications de l'expression des gènes. Bien que ces variations soient connues, il était difficilement envisageable de faire des échantillonnages de nuit dans les habitations, alors que cette période était propice pour la capture des individus de *T. brasiliensis* en milieu sylvatique. Le piégeage lumineux des *R. robustus* de terrain se justifiait par le fait que l'échantillonnage des *Rhodnius* dans les palmiers est difficile à mettre en œuvre en milieu forestier dense car il requiert soit de placer des pièges dans les palmiers (mais cette technique n'est pas toujours efficace et les palmiers sont parfois inaccessibles), soit d'abattre les palmiers ce qui n'est pas envisageable en forêt dense. La capture manuelle quand elle était possible a été privilégiée (pour *T. brasiliensis*) car elle permet de connaître précisément la zone géographique d'où viennent les vecteurs. De plus, certaines espèces du complexe *T. brasiliensis* également ciblées dans l'échantillonnage pour des études épidémiologiques se déplacent préférentiellement en marchant et certaines populations ou espèces sont brachyptères comme *T. sherloki* (Almeida et al., 2012)

Cependant, les recoupements qui ont pu être faits entre nos différentes analyses suggèrent une relative fiabilité de nos résultats. En effet, à travers les analyses d'expression différentielle, et malgré des procédés d'échantillonnage différents, des rapprochements entre nos deux modèles *Rhodnius* et *T. brasiliensis* peuvent être mis en évidence, suggérant des mécanismes de régulation des gènes en réponse à la domiciliation communs pour les deux modèles. D'un point de vue méthodologique, bien que les logiciels d'analyse d'expression différentielle EdgeR et DESeq2 n'estiment pas le même nombre de contigs différentiellement exprimés (DE), ils s'accordent sur une part importante de ces contigs. C'est à partir de ces similitudes que nous avons basé notre analyse pour le modèle *T. brasiliensis*. De plus, le

regroupement des contigs en fonction de leur dynamique d'expression par HTScluster confirme le comportement des contigs déterminés comme DE par DESeq2 et EdgeR. De la même manière, plusieurs éléments (expression différentielle mais aussi étude de présence et absence de transcrits) montrent que les résultats obtenus sont cohérents en ce qui concerne l'analyse d'expression des espèces *R. prolixus* et *R. robustus*.

Utilisation du modèle *Triatominae* pour l'évaluation de différentes méthodes d'assemblage RNAseq et d'analyse d'expression.

Assemblage transcriptomique

Mon premier travail a porté sur la réalisation de transcriptomes de référence les plus complets possibles pour nos espèces modèles. Pour cela, nous avons comparé différents procédés d'assemblage, basés sur diverses sources de données et de séquençage. Notre étude a montré que la réalisation d'un assemblage de qualité dépendait de divers paramètres, dont la procédure d'échantillonnage, la stratégie de séquençage et la stratégie d'assemblage utilisée. La variabilité génétique conduisant à des ambiguïtés et rendant l'assemblage du transcriptome par les logiciels difficile (Martin and Wang, 2011), les différences de qualité d'assemblage entre nos modèles peuvent être imputées au taux de polymorphisme variable d'une espèce à l'autre. Le nombre d'individus par échantillon ainsi que la provenance des échantillons (élevage ou terrain) vont impacter la variabilité au sein des transcriptomes et par conséquent jouer sur la qualité des assemblages. De plus, les difficultés rencontrées pour l'assemblage du transcriptome de *T. brasiliensis* pourraient également être indirectement corrélées à la complexité de son génome. En effet, la taille du génome de *T. brasiliensis* est estimée à 1Gb, (alors qu'elle n'est que de 670Mb chez *R. prolixus*) et refléterait des événements supplémentaires de duplication dans cette lignée évolutive. Si ces duplications comprennent des zones codantes (familles multigéniques), elles vont également impacter le transcriptome et complexifier son assemblage. Par ailleurs, l'extraction de l'ARN des échantillons des deux espèces sœurs *R. prolixus* et *R. robustus* n'a pas été réalisée simultanément avec celle des autres espèces de *Rhodnius* ni séquencée par la même plateforme. Il en a résulté une profondeur de séquençage inférieure à celle des autres *Rhodnius* (20 à 42M de paires de reads pour *R. prolixus* et *R. robustus* contre 99 à 152M de paires de reads pour les autres *Rhodnius*). Or, Francis et al., (2013) ont montré que la profondeur de séquençage impactait considérablement les résultats des assemblages *de novo*, et ce sur l'ensemble des espèces animales non-modèles étudiées et avec les deux logiciels utilisés, Oases et Trinity. C'est pourquoi ils recommandent une profondeur de séquençage comprise entre 50 et 60M de paires de reads pour optimiser la longueur des contigs et la complétude du transcriptome. Ces différents facteurs expliquent que les échantillons de *R. prolixus* et *R. robustus* présentaient une qualité d'assemblage *de novo* très inférieure à celle des 8 autres échantillons de *Rhodnius*. Au vu de nos résultats, nous préconisons de réduire au maximum le polymorphisme intrinsèque de l'échantillon à séquencer et de favoriser la

technologie paired-end au single-end avec un taux de couverture suffisamment important pour la réalisation d'un assemblage *de novo*.

Par ailleurs, nous avons pu montrer que l'utilisation d'une stratégie de séquençage mixte qui combine à la fois des techniques générant des longs reads (454 dans notre cas) et des techniques favorisant une couverture élevée de séquençage (Illumina) nous a permis d'améliorer considérablement nos assemblages *de novo*. Des études récentes sur d'autres organismes non-modèles ont mené à des conclusions similaires comme par exemple l'assemblage du transcriptome de la caille japonaise par Finseth et Harrison (2014). D'autre part, Ibarra-Laclette et al., (2015) se sont basés sur le workflow que nous avons publié dans Marchant et al., (2015) pour réaliser l'assemblage *de novo* de l'avocat *Persea americana*. La comparaison des transcrits ainsi assemblés avec le génome de cette espèce a montré un succès significatif (94,65% des transcrits alignés sur plus de 98% de leur longueur avec au minimum une identité de 90%). Par ailleurs, aujourd'hui une 3^{ème} génération de séquençage voit le jour. La technologie Pacific Biosciences est actuellement le leader de cette dernière génération. Elle permet de séquencer des fragments d'ADN de très grande taille (jusqu'à 2500 nucléotides) à partir d'une seule molécule d'ADN. Cependant, cette technologie génère de nombreuses erreurs de séquençage. Mais de la même manière que le 454, ces erreurs peuvent être corrigées par l'utilisation de données Illumina. Ainsi, la combinaison des technologies PacBio et Illumina se révèle avantageuse dans les problématiques d'assemblages *de novo* (Koren et al., 2012).

Quand un génome de référence est disponible, les assemblages basés sur cette référence sont considérés comme plus efficaces que l'assemblage *de novo* (Martin and Wang, 2011). Cette stratégie a donné de bien meilleurs résultats pour les échantillons *R. prolixus* et *R. robustus* que ceux obtenus *via* les assemblages *de novo* pour ces deux espèces. En effet, contrairement à l'assemblage *de novo* qui nécessite un taux de recouvrement important pour assembler plusieurs reads en un seul contig, l'utilisation d'une référence permet d'assembler des transcrits en plus faible abondance (Martin and Wang, 2011). Le transcriptome de ces deux espèces possédant une couverture relativement faible, cet avantage n'est donc pas négligeable. Par ailleurs, les très bons résultats issus de l'assemblage de *R. robustus* basé sur le génome de *R. prolixus* confirme qu'en absence d'un génome de référence, il est possible d'utiliser la référence d'une espèce étroitement liée (Hornett and Wheat, 2012; Ward et al., 2012).

Choisir entre la réalisation d'un assemblage *de novo* ou basé sur une référence dépend aussi largement de l'utilisation finale du transcriptome de référence (Martin and Wang, 2011). En effet, pour une analyse d'expression différentielle, il est important d'obtenir un transcriptome le moins fragmenté possible, avec le moins de répétitions possible afin de ne pas biaiser les données de comptage utilisées pour l'analyse. De plus, seule la présence des transcrits partagés par les différents échantillons est nécessaire pour une telle analyse. Ainsi, nous avons porté notre choix sur un assemblage basé sur un génome de référence pour l'analyse d'expression différentielle entre les deux espèces sœurs *R. prolixus* et *R. robustus* afin d'optimiser la longueur assemblée des transcrits. En revanche, pour l'annotation et la

comparaison de familles multigéniques, l'assemblage basé sur une référence peut présenter des inconvénients. En effet, lorsque l'assemblage se base sur le génome d'une autre espèce étroitement liée au modèle étudié, il ne tient pas compte des transcrits provenant de régions génomiques divergentes entre le modèle étudié et la référence (Ward et al., 2012). C'est pourquoi nous avons privilégié les assemblages *de novo* pour réaliser l'annotation des OBPs et des CSPs des différentes espèces de *Rhodnius*.

Analyse d'expression différentielle

Les nouvelles technologies de séquençage sont apparues dans les années 2000 et ont permis l'élaboration du RNAseq, c'est-à-dire le séquençage à haut débit de données transcriptomiques. Pour exploiter les possibilités et relever les défis posés par ce type de données relativement nouveau, un certain nombre de logiciels ont été développés spécialement pour l'analyse différentielle de l'expression des gènes. Cependant, il y a un manque de consensus sur la stratégie optimale d'analyse et sur le choix des logiciels appropriés à utiliser. Un certain nombre de paramètres influencent l'analyse à savoir le nombre de réplicats, la profondeur de séquençage, la divergence entre les réplicats et l'équilibre de la taille de librairie des différents échantillons. Le choix des logiciels à utiliser dépend en partie de ces paramètres (Seyednasrollah et al., 2015 ; Zhang et al., 2014b). Seyednasrollah et al., (2015) ont comparé plusieurs logiciels en étudiant d'une part les différences d'expression entre deux souches de souris d'origine différente et d'autre part les différences d'expression en fonction du sexe sur des échantillons de lymphoblastes humains. Au sein d'une même souche de souris, l'expression est homogène tandis qu'au sein d'un même sexe pour le modèle humain, la variabilité est bien plus élevée. Ils ont montré que les résultats pouvaient être très différents entre les pipelines en fonction des modèles, de leur polymorphisme et du nombre de réplicats utilisés. Ainsi, il n'y aurait pas de méthode unique susceptible d'être optimale en toutes circonstances. Ces auteurs conseillent d'exécuter les analyses avec plusieurs logiciels afin de renforcer la validité des résultats obtenus. Nous avons choisi d'utiliser les deux packages R DESeq2 et EdgeR qui sont le plus souvent recommandés. DESeq2 est considéré comme étant plus performant en cas de petit nombre de réplicats ou de forte divergence entre échantillons (Seyednasrollah et al., 2015) mais il présente le défaut d'être sensible aux différences de tailles de librairies entre les échantillons (Zhang et al., 2014b). EdgeR génère quelques faux positifs mais est plus stable que DESeq2 en cas de tailles de librairies très divergentes et supporte mieux des plans d'expériences complexes (Seyednasrollah et al., 2015 ; Zhang et al., 2014b). La méthode Cuffdiff2 présente l'avantage d'être incluse dans le logiciel Cufflink qui nous a permis de générer l'assemblage des échantillons *R. prolixus* et *R. robustus*. Néanmoins, pour l'analyse d'expression différentielle, ce logiciel est très souvent critiqué. Il nécessite un temps de calcul bien plus important et détecte bien moins de gènes DE par rapport aux deux packages R précédemment évoqués (Seyednasrollah et al., 2015; Zhang et al., 2014b). C'est pourquoi nous avons choisi de

ne pas utiliser cette méthode pour notre analyse d'expression différentielle entre les deux espèces de *Rhodnius*.

Le logiciel DESeq2 a évalué un nombre de contigs différentiellement exprimés plus important que le logiciel EdgeR, et ceci dans les deux analyses (*T. brasiliensis* et *Rhodnius*). Cependant, en ce qui concerne l'analyse chez *T. brasiliensis*, près de 75% des transcrits sont filtrés par DESeq2 du fait de leur faible expression. Ce filtrage en masse implique beaucoup moins de tests d'hypothèses effectués par DESeq2 et donc une correction pour les tests multiples bien moins stricte. Par conséquent, avec notre jeu de données, DESeq2 a détecté davantage de gènes différentiellement exprimés que EdgeR, allant à l'encontre des observations menées par Seyednasrollah et al., (2015) et par Zhang et al., (2014b). Nos échantillons de *T. brasiliensis* présentent des tailles de bibliothèques très déséquilibrées (22 à 71M de reads en excluant PMD), ce déséquilibre étant renforcé par l'absence de réplicats techniques pour certains échantillons (22 à 94M de reads après réunion des réplicats techniques). DESeq2 étant sensible à ce déséquilibre, il est possible que l'importante part de données filtrées soit liée à des différences de tailles de bibliothèques de nos échantillons. Nous avons tout de même considéré que les contigs évalués comme différentiellement exprimés par les deux logiciels, EdgeR et DESeq2 devaient être plus fiables que ceux uniquement détectés par EdgeR.

EdgeR a identifié très peu de contigs différentiellement exprimés lors de la comparaison *R. prolixus*/*R. robustus*, réduisant très fortement les possibilités d'annotation de nos données DE. L'échantillon RMG étant très divergent des autres échantillons, il est possible que cette forte hétérogénéité au sein de nos échantillons ait pu diminuer les capacités de ce logiciel. DESeq2 est considéré comme plus robuste dans de telles conditions. Cependant, à l'inverse de nos résultats, pour le modèle humain considéré comme possédant une variabilité importante entre réplicats dans l'étude de Seyednasrollah et al., (2015) EdgeR dénombre plus de gènes DE (dont des faux positifs) que DESeq. Cependant, Seyednasrollah et al., (2015) utilisent DESeq et non DESeq2, ce qui pourrait expliquer cette différence. Ching et al., (2014) ont comparé EdgeR, DESeq et DESeq2 et trouvent en effet que EdgeR présente plus de gènes DE que DESeq. Ils trouvent aussi que DESeq2 et EdgeR montrent un nombre de gènes identifiés comme DE globalement équivalents. Cependant, pour l'étude sur les deux espèces sœurs de *Rhodnius*, le déséquilibre de taille des bibliothèques entre échantillons était bien moins marqué (20 à 32M de reads) que dans l'étude portant sur *T. brasiliensis*, ce qui conduit à la filtration de seulement 0,57% des données par DESeq2. Les résultats de ce logiciel pour l'analyse *R. prolixus*/*R. robustus* semblent donc plus fiables que pour l'analyse précédente. C'est pourquoi, nous avons choisi de nous baser sur DESeq2 pour l'analyse des résultats *R. prolixus*/*R. robustus*.

Implication de l'expression des transcrits des organes chimiosensoriels dans l'adaptation aux anthroposystèmes chez les Triatominae

Cette étude avait pour objectif principal d'aborder la problématique liée à la domiciliation des vecteurs de la maladie de Chagas et à ses mécanismes adaptatifs en se basant sur des comparaisons d'expression des gènes issus des organes chimiosensoriels de Triatominae. Les analyses d'expression différentielle entre les échantillons sylvatiques et domiciliés réalisées sur le modèle *T. brasiliensis* et celles réalisées sur les deux espèces sœurs *R. prolixus* et *R. robustus* ont révélé des similitudes de dynamique d'expression entre les deux environnements, mais aussi des divergences.

Les échantillons de *T. brasiliensis* domiciliés présentaient une sous-expression globale des contigs DE en comparaison des échantillons péri-domiciliés et sylvatiques. Cependant, ce déséquilibre n'est pas retrouvé entre les échantillons d'élevage et les échantillons sylvatiques de *Rhodnius* où on relève un nombre de gènes sur-exprimés et sous-exprimés équivalent dans chacune des conditions. Il est donc difficile de dégager une tendance sur les données globales. En revanche, plusieurs familles multigéniques ont montré des divergences d'expression similaires entre les vecteurs domiciliés et les sylvatiques des deux modèles. C'est notamment le cas de familles multigéniques impliquées dans la protection contre les stress engendrés par l'environnement : les cytochromes P450 et les glutathione S-transférase, impliqués dans la détoxification ; les heat-shock protéines, jouant un rôle dans la résistance aux températures extrêmes et à la dessiccation, et les protéines cuticulaires, protégeant plus globalement les insectes du milieu extérieur (par exemple en empêchant l'entrée d'éléments étrangers ou en jouant le rôle de barrière à la perte d'eau). La plupart des gènes appartenant à ces familles multigéniques sont sous-exprimés dans les échantillons domiciliés ou d'élevage par rapport aux échantillons sylvatiques, quel que soit le modèle. L'environnement d'élevage comme les habitations humaines présente beaucoup moins de variations biotiques et abiotiques que l'environnement sauvage ; il se pourrait que ce milieu engendre moins d'agressions extérieures, diminuant la réponse des punaises et donc l'expression de gènes de protection. Il est intéressant de noter que des résultats similaires ont été retrouvés dans l'étude de Glaser et al., (2015) sur des populations de noctuelles. Les individus d'une population kenyane de *S. nonagrioides* échantillonnés dans un environnement sauvage surexpriment des gènes codant pour des cytochromes P450, des glutathione S-transférase et des estérases par rapport à une population française qui évolue en milieu anthropisé. Cependant, de nombreux insectes surexpriment ces gènes de décontamination quand ils sont en contact avec des environnements anthropisés traités avec des insecticides ou pollués. C'est le cas d'*A. aegypti*, *A. albopictus*, *An. gambiae* (David et al., 2010; Nkya et al., 2013), *A. mellifera* (Gregorc et al., 2012), *D. melanogaster* (Daborn et al., 2007; Harrop et al., 2014), du puceron *Myzus persicae* (Puinean et al., 2010), ou encore de la punaise de lit *Cimex lectularius* (Mamidala et al., 2012; Zhu et al., 2013). La régulation positive des gènes de détoxification a été mise en lien avec l'adaptation aux insecticides et donc la résistance à ces derniers (Roush and Tabashnik, 2012). Cependant, bien

que nous ne contrôlions pas l'utilisation des insecticides ménagers par les habitants, les domiciles dans lesquels l'échantillonnage des *T. brasiliensis* a été réalisé n'avaient pas subi de campagnes d'insecticides nationales récentes. Il est donc logique de ne pas retrouver de surexpression des gènes de détoxification au sein des échantillons de *T. brasiliensis* domiciliés.

Des contigs annotés comme appartenant à la famille multigénique des *takeout* se sont révélés sous-exprimés au sein des individus *T. brasiliensis* domiciliés par rapport aux sylvatiques. En revanche, cette observation ne semble pas généralisable au sein des Triatominae puisqu'aucune différence d'expression impliquant la famille des *takeout* n'a pu être observée pour *Rhodnius*. Les protéines *takeout* sont impliquées dans le cycle circadien, dans le métabolisme de l'alimentation, dans le comportement de recherche alimentaire (Sarov-Blat et al., 2000; So et al., 2000) et dans la régulation de l'activité locomotrice pendant l'activité de recherche de nourriture (Meunier et al., 2007). D'autres études d'analyse d'expression différentielle ont pu mettre en évidence l'implication de protéines *takeout* dans des différences de comportement chez les insectes. Par exemple, *Danaus plexippus* (papillon monarque d'Amérique du Nord) est une espèce qui migre en automne. Cette espèce montre une sous-expression des protéines *takeout* chez les migrants par rapport aux papillons d'été sédentaires. Ce changement d'expression des *takeout* pourrait être en lien avec l'utilisation des changements de durée d'ensoleillement par les migrants pour les aider à naviguer vers leurs aires d'hivernage (Zhu et al., 2008). L'implication de cette famille multigénique dans les modifications comportementales de criquets migratoires *Locusta migratoria* a également été mise en évidence. Ces derniers changent de comportement en réponse aux changements de densité de population, ils peuvent passer ainsi d'une phase grégaire à une phase solitaire de migration. Guo et al., (2011) ont mis en évidence des gènes, appartenant à plusieurs catégories en rapport avec la perception olfactive périphérique, différentiellement exprimés entre des criquets issus de ces deux phases. Parmi ces gènes, plusieurs CSPs ont montré une expression plus élevée chez les criquets grégaires, alors qu'à l'inverse, le gène *takeout* LmigTO1 était sur-exprimé chez les solitaires. Des expériences comportementales olfactives combinées à des méthodes de silencing ont montré que les gènes de CSP DE étaient impliqués dans l'attraction entre les individus au cours de la grégarisation alors que l'expression de la LmigTO1 entraînait la répulsion des individus dans la phase de migration. Les échantillonnages de *T. brasiliensis* sylvatiques ont très probablement impliqué la récolte d'insectes en cours de dispersion. Les punaises domiciliées quant à elles ont été échantillonnées directement dans les fissures des murs où elles avaient instauré leurs colonies, soit sous forme agrégée. Au vu des résultats de Guo et al. (2011) et de notre échantillonnage, il est envisageable que la sur-expression des gènes *takeout* dans les échantillons de *T. brasiliensis* sylvatiques par rapport aux domiciliés soit liée à des différences de comportement de dispersion *versus* agrégation. L'expression des gènes de *takeout* est également induite par la faim (Meunier et al., 2007). La sous-expression de ces gènes au sein des échantillons domiciliés pourrait donc être expliquée par un état de satiété plus important dans la population domiciliée.

Les gènes chimiosensoriels ne montrent pas tous les mêmes variations d'expression entre environnements chez *T. brasiliensis* et chez *Rhodnius*. Cependant, certaines OBPs et CSPs ont révélé pour les deux modèles une sous-expression dans les échantillons d'environnement anthropisé (élevage ou domicilié). De plus, sur les trois CSPs retrouvées sous-exprimées dans les échantillons de *Rhodnius* d'élevage, les CSP1 et 11 possèdent des orthologues également sous-exprimés dans les échantillons domiciliés de *T. brasiliensis*. Cette même observation peut être faite pour l'OBP23. Cependant, certains résultats divergent également entre les deux taxons. Peu d'OBPs et de CSPs ont été identifiées comme différentiellement exprimées entre les échantillons de *Rhodnius* d'élevage et sylvatiques par rapport aux résultats obtenus sur le modèle *T. brasiliensis*, différence qui pourrait être attribuée à une spécificité des modèles. Le modèle *Rhodnius* permet de détecter également des récepteurs chimiosensoriels (ORs, IRs et GRs) différentiellement exprimés, essentiellement sous-exprimés chez les *Rhodnius* d'élevage. L'absence de récepteurs chimiosensoriels différentiellement exprimés au sein du modèle *T. brasiliensis* peut s'expliquer par le fait qu'ils ne sont pas retrouvés dans le transcriptome de référence, ce qui est probablement dû à leur expression relativement faible, rendant leur assemblage difficile via des méthodes *de novo*. L'utilisation de l'assemblage par référence a probablement aidé à leur détection pour le modèle *Rhodnius*.

Le passage d'un milieu sauvage à un milieu anthropisé implique des changements d'hôtes pour ces deux modèles. Ces différences d'expression entre les échantillons provenant d'environnements différents pourraient donc être imputées à l'adaptation de ces insectes à différents hôtes. Des études antérieures ont pu mettre en évidence des différences d'expression de gènes chimiosensoriels entre des espèces d'insectes ou des populations adaptées à des hôtes différents. C'est le cas de deux espèces de noctuelles, l'une adaptée au coton (*Helicoverpa armigera*), l'autre au tabac (*H. assulta*), et qui diffèrent par l'expression d'une OBP (Li et al., 2013). Chez *Sesamia nonagrioides*, l'analyse d'expression différentielle entre une population kenyane récoltée sur des plants de Poacées sauvages (*Typha domingensis*) et une population française récoltée dans des cultures de maïs avait également révélé des différences d'expression au sein des protéines de liaisons olfactives (Glaser et al., 2015). Se rapprochant davantage de nos problématiques, McBride et al., (2014) ont travaillé sur le moustique *Aedes aegypti* qui présente deux sous espèces : *Ae. aegypti formosus* retrouvée dans des milieux sylvatiques et qui se nourrit sur des animaux, et *Ae. aegypti aegypti*, anthropophile. Ils ont montré que l'évolution de la préférence pour l'odeur humaine chez les moustiques anthropophiles était étroitement liée à l'augmentation de l'expression du récepteur olfactif AegOr4 qui lie une molécule présente à des niveaux élevés dans l'odeur humaine.

Compte tenu de la convergence de nos résultats pour nos deux modèles, et au regard d'autres études similaires, l'adaptation des vecteurs de la maladie de Chagas à l'environnement anthropique semble liée à des différences de sensibilité olfactive, confortant notre hypothèse de départ. Ces différences de sensibilité olfactive peuvent résulter d'une part de la régulation de l'expression génique des OBPs et des CSPs, ce qui module l'apport des substances odorantes

aux récepteurs, et d'autre part de la régulation de l'expression des ORs aboutissant à la modulation de leur réception. Cependant, nous ne disposons pas de données fonctionnelles pour les gènes chimiosensoriels exprimés différemment chez nos modèles *Triatominae*, leur rôle olfactif reste donc spéculatif.

Plasticité phénotypique ou adaptation génique ?

La plasticité phénotypique est la capacité d'un organisme à exprimer différents phénotypes à partir d'un génotype donné selon les conditions environnementales (Agrawal, 2001). Ainsi, un génotype peut s'exprimer différemment selon l'environnement. Dans ce cas, la plasticité phénotypique d'un organisme peut lui permettre de s'adapter à différents environnements sans passer obligatoirement par un changement génétique. Les réponses à des changements environnementaux par la plasticité phénotypique impliquent des changements de l'expression des gènes. Des études révèlent que dans certains cas, jusqu'à 50% du génome peut avoir une expression variable en fonction de l'environnement (Snell-Rood et al., 2010). Dans notre étude, l'observation de gènes différemment exprimés entre les individus de *Triatominae* sylvatiques et domiciliés est plus probablement liée à de la plasticité phénotypique.

Un problème de longue date en biologie évolutive est de comprendre la relation entre la variation induite par l'environnement (observable sur une génération), et les changements évolutifs basés sur des différences génétiques hérissables entre générations. Dans les modèles traditionnels, la plasticité induite par l'environnement est considérée comme une variation non hérissable. Les modèles actuels reconnaissent que les environnements peuvent induire de la plasticité considérée comme adaptative ou non adaptative par rapport à un phénotype optimum local. La plasticité est considérée comme adaptative quand le phénotype est modifié dans la même direction que celle favorisée par la sélection naturelle. En revanche, la plasticité est considérée comme non-adaptative quand une population se heurte à un environnement qui induit la production de phénotypes qui s'écartent de l'optimum local, entraînant une corrélation négative entre la direction de la plasticité et la direction de la sélection. La plasticité peut cependant présenter un coût et ainsi réduire la valeur sélective des individus. De ce fait, si la plasticité adaptative présente un coût trop élevé, est incomplète, ou si la plasticité est non-adaptative ; alors l'adaptation génétique par sélection sera bénéfique pour la population. Des expériences de sélection en laboratoire ont en effet montré une relation positive entre le sens de l'évolution et les réponses adaptatives de la plasticité alors que cette relation est négative avec les réponses non adaptatives (Ghalambor et al., 2007 ; Ghalambor et al., 2015).

D'autre part, la plasticité peut être elle-même déterminée génétiquement et être variable entre individus. Elle est ainsi soumise comme d'autres caractères à la sélection qui peut favoriser des phénotypes divergents selon l'environnement. Par conséquent, la plasticité phénotypique est parfois à l'origine de la divergence entre les populations situées dans des

conditions environnementales distinctes (Pfennig et al., 2010). Plasticité adaptative et adaptation génique (sélection) peuvent donc agir conjointement pour la sélection d'un phénotype optimum.

R. prolixus et *R. robustus* ont montré des liens phylogénétiques très étroits dans notre analyse phylogénétique, révélant que leur statut spécifique n'est pas clair. Notre analyse d'expression différentielle entre les deux espèces sœurs *R. prolixus* et *R. robustus* a montré qu'il n'existait pas de différence d'expression entre ces deux espèces lorsqu'elles étaient toutes deux élevées dans les mêmes conditions. En revanche, les échantillons de *R. robustus* sylvatiques montrent d'importantes différences d'expression avec les *Rhodnius* d'élevage. *R. robustus* serait donc capable de modifier l'expression de ses gènes lorsqu'il est placé dans des conditions environnementales qui ne lui sont pas naturelles. Il présente alors un profil d'expression similaire à celui de *R. prolixus*, ce qui laisse supposer que la plasticité dont *R. robustus* fait preuve aurait également pu s'exprimer chez *R. prolixus* au moment de ses premières colonisations du milieu anthropique. Il est alors possible qu'une sélection divergente se soit exercée entre populations localisées dans des environnements différents menant jusqu'à l'isolement reproductif et donc à l'apparition d'un phénomène de spéciation (Pfennig et al., 2010).

Concernant *T. brasiliensis*, des flux géniques ont été mis en évidence par l'utilisation de marqueurs microsatellites entre les différentes populations notamment entre la population domiciliée B et la population péri-domiciliée D, ainsi qu'entre la population sylvatique A et la population péri-domiciliée D (Almeida et al., soumis). Ce brassage génétique pourrait donc limiter la structuration génétique des populations et par conséquent la mise en place de processus adaptatifs par ségrégation. La récente domiciliation de *T. brasiliensis* ainsi que l'importance des flux migratoires entre les populations sylvatiques et domiciliées nous poussent, dans un premier temps, à privilégier l'hypothèse de la plasticité plutôt que celle d'une évolution génétique spécifique des populations domiciliées pour cette espèce. Il est probable que l'instabilité environnementale récente subie par les populations de *T. brasiliensis* (déforestation, changement climatique, diminution de la diversité des hôtes...) ait mis en évidence la plasticité, permettant aux individus de s'ajuster au mieux à ces variations et de se développer dans de nouveaux environnements, notamment les anthroposystèmes. L'implication de la plasticité phénotypique chez *T. brasiliensis* a déjà été proposée en réponse à des changements environnementaux. Batista et al., (2013), ont observé des différences morphologiques entre les ailes de populations de *T. brasiliensis* sylvatiques et celles de populations péri-domiciliées et supposent que ces différences sont liées à l'existence de la plasticité phénotypique associées à différents écotopes, possiblement en raison de la situation dans chaque micro-habitat, telles que la température, l'humidité relative, la disponibilité en ressources alimentaires et la densité des populations.

Il n'est également pas à exclure que les populations domiciliées de *T. brasiliensis* puissent un jour diverger des populations sylvatiques menant à une structuration plus drastique

voir même à l'apparition de deux espèces distinctes. Fitzpatrick et al., (2008) et Thibert-Plante and Hendry, (2011) soulignent que l'existence d'une interaction entre dispersion et plasticité peut accentuer ou au contraire limiter la divergence entre deux populations. Par exemple, si la dispersion d'un individu s'effectue au stade adulte, ses traits plastiques seront fixés avant la dispersion. Ainsi, parce que les conditions qui déterminent le phénotype sont celles expérimentées dans l'habitat natal, si un adulte se disperse et s'installe dans un nouvel habitat, le phénotype exprimé peut être mal adapté à ce nouvel environnement. La valeur sélective de cet adulte nouvellement installé sera donc inférieure à celle des individus nés dans ce milieu, limitant alors le flux de gènes entre la population autochtone et le nouvel arrivant. Donc, malgré les phénomènes de migration qui ont pu être mis en évidence entre les populations sylvatiques et domiciliées chez *T. brasiliensis* (Almeida et al., soumis), il n'est pas impossible que les populations divergent au point de se scinder en deux espèces différentes. De plus, les populations étant fondées par un petit groupe de migrants, l'effet de la dérive génétique n'est pas à négliger dans un processus de spéciation. Enfin, le relâchement de la pression du contrôle chimique déjà noté dans certains pays comme au Brésil, ou l'apparition de résistance aux insecticides, pourrait accentuer les processus de spéciation.

L'épigénétique participe à la plasticité phénotypique en induisant des changements d'activité des gènes, donc des changements de caractères, qui sont transmis au fil des divisions cellulaires ou des générations sans faire appel à des mutations de l'ADN (Feil and Fraga, 2012). Les différences d'expression relevées entre les Triatominae originaires d'habitats différents pourraient donc également être le reflet de modifications épigénétiques en réponse à l'environnement anthropique et pourraient mener à une adaptation à l'environnement domicilié de manière héritable mais réversible. Etudier les marques épigénétiques aux alentours des gènes d'intérêt différentiellement exprimés pourrait permettre de répondre à cette hypothèse.

Evolution des répertoires de gènes chimiosensoriels chez les Triatominae

L'expansion dynamique des gènes chimiosensoriels peut représenter une source importante de variation génétique et serait susceptible de modifier les caractères de reproduction, écologiques ou comportementaux des insectes. Ce processus offre donc des possibilités d'adaptations génétiques spécifiques (Sánchez-Gracia et al., 2009). La mise en place des familles multigéniques résulte de la duplication de gènes en deux copies évoluant chacune individuellement. La répétition de ce processus peut conduire à des familles constituées de dizaines voire de centaines de gènes, d'où la notion d'expansion de gènes. Suite à une duplication, une copie peut évoluer en pseudogène, en un gène différemment régulé (mutation dans la région régulatrice), ou être modifiée pour réaliser une nouvelle fonction (mutation dans la région codante) (Ohta, 2013). Ce mode de régulation dynamique au sein des gènes chimiosensoriels permettrait aux espèces de se spécialiser dans un environnement particulier (Guo and Kim, 2007). L'expansion des gènes chimiosensoriels impliquée dans l'adaptation à un environnement spécifique ou à un nouvel hôte a été mise en évidence plusieurs fois chez

différents insectes, notamment chez les drosophiles et les moustiques. Pour d'autres espèces (comme *A. mellifera*, *A. pisum* ou chez diverses espèces de fourmis), la spécialisation fonctionnelle n'a pas été montrée directement mais a été suggérée fortement par l'observation de traces de sélection positive. L'étude présentée ici n'a pas porté sur la recherche de traces de sélection positive ou sur le recrutement fonctionnel de nouvelles copies. Ce travail a permis en revanche la comparaison du répertoire de gènes chimiosensoriels entre espèces proches et donc l'étude de la dynamique d'expansion des familles multigéniques concernées en fonction des conditions de vie. Le récent séquençage du génome de *Rhodnius prolixus* a permis l'annotation des gènes chimiosensoriels pour cette espèce (Mesquita et al., 2015). A partir de cette référence et des données transcriptomiques, nous avons pu évaluer le nombre de gènes appartenant à la famille multigénique des OBPs et des CSPs chez 9 autres espèces de *Rhodnius* et chez *T. brasiliensis*. Si l'on compare ces répertoires géniques à ceux d'autres Paraneoptera pour lesquels le génome est entièrement séquencé, on constate que le nombre d'OBPs et de CSPs chez les Triatominae est globalement supérieur à leurs homologues. Les autres gènes chimiosensoriels (ORs, GRs, IRs) étant généralement faiblement exprimés, il est difficile de faire des comparaisons avec des données transcriptomiques. Du fait de la difficulté à assembler *de novo* les gènes faiblement exprimés, l'ensemble des récepteurs chimiosensoriels des Triatominae n'ont pas été annotés dans cette étude. Par ailleurs, les génomes des différentes espèces du genre *Rhodnius* dont nous disposons étant très morcelés, il est difficile d'y annoter les gènes sans s'aider des transcrits, c'est pourquoi cette analyse n'a pas été réalisée. Mais la comparaison avec les autres Paraneoptera pour ces répertoires peut être évaluée avec *R. prolixus*. On retrouve chez cette espèce une expansion des gènes annotés comme ORs potentiels.

Les Triatominae étant associés à des habitats très divers (Gaunt et Miles, 2000) et à un large éventail d'hôtes vertébrés (Patterson, 1994), l'expansion des OBPs, des CSPs et des ORs pourrait être liée à leur adaptation en tant qu'insecte hématophage à une diversité importante de niches écologiques.

Au sein même du genre *Rhodnius*, ces événements d'expansion ont été mis en relation avec l'habitat des différentes espèces. Ainsi, nous avons montré que le nombre de gènes d'OBPs et de transcrits OBPs et CSPs était corrélé positivement à la domiciliation des espèces. Par ailleurs, les analyses ont montré un nombre de transcrits chimiosensoriels détectés plus important dans les échantillons d'élevage que dans les échantillons sauvages. Ce résultat suggère que la corrélation positive entre domiciliation et le nombre de transcrits chimiosensoriels pourrait être retrouvée entre les populations domiciliées et sylvatiques des espèces de *Rhodnius* qui colonisent plusieurs écotopes. La capacité à coloniser l'habitat domestique par certaines espèces de Triatominae pourrait donc être liée à des différences au niveau des familles multigéniques chimiosensorielles. Une comparaison approfondie des séquences codantes et régulatrices de ces différents groupes de gènes sur plusieurs espèces de Triatominae, notamment chez les *Rhodnius*, ainsi qu'une étude des traces de sélection dans ces gènes et dans leur voisinage pourraient nous aider à comprendre les liens entre l'expression des

gènes chimiosensoriels chez les punaises hématophages, l'évolution du répertoire de ces gènes et les pressions de sélection exercées par leur environnement.

Par ailleurs, nos résultats de corrélation suggèrent une expansion des gènes OBPs et CSPs et une plus grande proportion de leurs transcrits exprimés chez les *Rhodnius* en lien avec leur capacité à la domiciliation. D'autre part, les analyses d'expression de gènes ont révélé une sous-expression des gènes chimiosensoriels chez les individus Triatominae domiciliés. Des études ont montré chez la souris et la drosophile que certains gènes présentaient un niveau d'expression négativement corrélé avec le nombre de duplications de ces gènes (Dopman and Hartl, 2007; Henrichsen et al., 2009a). Le mécanisme sous-jacent est encore mal compris, mais peut être expliqué par deux modèles. Dans le premier modèle, une corrélation négative entre le nombre de copies et l'expression relative est expliquée par l'apparition de gènes précoces. Ces gènes sont d'abord exprimés à des niveaux proportionnels à leur nombre de copies puis ils induisent directement ou indirectement l'expression d'un répresseur qui, par une boucle de rétro-action, réduit l'expression des autres copies de ce gène. Dans la seconde hypothèse, les copies supplémentaires d'un gène par encombrement stérique altèrent l'accès de ce locus à la machinerie cellulaire de transcription (Henrichsen et al., 2009b). Ainsi, on peut supposer que l'expansion des gènes chimiosensoriels en lien avec l'adaptation des Triatominae aux anthroposystèmes aurait impliqué une baisse de l'expression de certains gènes dupliqués.

D'autre part, la capacité d'une espèce à générer plusieurs isoformes à partir d'un même gène permet également des variations dans l'utilisation de ce gène. Ce processus peut mener à l'apparition de nouvelles fonctions. Par exemple le récepteur gustatif GR39a chez *Drosophila melanogaster* possède quatre grands exons qui sont épissés alternativement avec trois petits exons conservés (Clyne et al., 2000). Ce récepteur est impliqué dans le comportement de cour chez les mâles, probablement par la réception d'une phéromone (Watanabe et al., 2011). Les orthologues du gène Gr39a ont été identifiés dans les génomes séquencés de 11 autres espèces de drosophiles. Leur comparaison a montré une évolution inhabituelle de la structure de ce récepteur par rapport aux autres. Gardiner et al., (2008) ont étudié en détail l'évolution du gène Gr39a et ont découvert des preuves de sélection positive qui, selon eux, suggèrent que la présence d'exons en tandem et leur épissage alternatif sont impliqués dans l'apparition de nouvelles fonctions, bien qu'elles restent actuellement inconnues. Ils ont montré que les estimations de divergence fonctionnelle entre les exons étaient d'une ampleur similaire à la divergence fonctionnelle qui existait entre les gènes dupliqués, ce qui suggère que la conséquence d'une divergence d'exon est largement équivalente à celle d'une duplication de gènes. Cependant, l'épissage alternatif est considéré comme rare au sein des récepteurs chimiosensoriels (Hill et al., 2002; Robertson and Wanner, 2006). Néanmoins, chez l'hémiptère *Lygus lineolaris*, l'épissage alternatif de gènes OBPs semble contribuer de manière significative à la multiplicité des séquences d'OBPs pour cette espèce (Hull et al., 2014). Nous avons également pu observer chez les Triatominae la présence de gènes chimiosensoriels pour lesquels plusieurs isoformes ont pu être identifiées. Ainsi, il est envisageable que l'épissage alternatif

d'un gène codant une OBP, notamment s'il a lieu au niveau du site actif, conduise à deux isoformes se liant à des molécules odorantes différentes. Ces divergences d'isoformes pourraient être impliquées dans la divergence d'habitat entre les différentes espèces de Triatominae.

Perspectives

Suite à nos différentes études, des perspectives peuvent être envisagées prenant en compte notamment des aspects méthodologiques mais également des analyses complémentaires à effectuer ainsi que des pistes de recherche visant le contrôle des Triatominae.

Elaboration d'un nouveau plan d'échantillonnage

Les technologies du séquençage à haut débit ayant progressé très rapidement ces dernières années, la quantité d'ARN nécessaire au séquençage est bien moins contraignante actuellement qu'au moment de l'élaboration de ce projet. De plus, les variations techniques liées à ce type de séquençage sont désormais contrôlées. Le recul sur les analyses et les collaborations avec les statisticiens nous ont permis de comprendre les priorités à accorder à un plan d'échantillonnage. Ainsi, dans le but de réaliser une analyse d'expression différentielle, on recommande à l'heure actuelle de privilégier les réplicats biologiques aux réplicats techniques et à la couverture de séquençage. En prenant en compte les capacités techniques actuelles ainsi que les contraintes de terrain et budgétaires, des recommandations peuvent être émises. Nous ne pouvons que recommander d'utiliser davantage de réplicats biologiques, notamment en incluant plusieurs populations domiciliées et peridomiciliées au sein de l'espèce *T. brasiliensis*, afin d'inclure les variations inter-populationnelles, étude que nous avons envisagée mais que la mauvaise conservation des échantillons ne nous a pas permis de faire. En ce qui concerne l'étude des *Rhodnius*, il serait très intéressant d'obtenir des échantillons de *R. prolixus* domiciliés prélevés sur le terrain (comparativement à des lignées d'élevage). Nous pourrions également utiliser des échantillons des quelques rares populations de *R. prolixus* sylvatiques reportées au Venezuela ou en Colombie. L'utilisation de plusieurs populations par espèce de *Rhodnius* permettrait également de généraliser les résultats. De plus, le transcriptome de ces espèces étant déjà réalisé, nous serions moins contraints par la couverture de séquençage nécessaire à l'assemblage *de novo*, ce qui permettrait de privilégier les réplicats biologiques à la couverture de séquençage pour une même enveloppe budgétaire. En revanche, le séquençage en Illumina paired-end, bien que plus coûteux, est vraiment un atout à ne pas négliger car il présente l'avantage d'engendrer moins d'erreurs lors de l'étape d'alignement des lectures sur la référence.

Mécanismes adaptatifs

La part de la plasticité et de l'adaptation génétique dans la réponse à un changement environnemental peut être appréhendée par la mise en place d'élevages à partir des différentes populations sylvatiques, péri-domiciliées et domiciliées. Ces élevages sont en cours au laboratoire de l'Insetário de Triatominae da Faculdade de Ciências Farmacéuticas, Universidade Estadual Paulista (UNESP) - Araraquara dans cet objectif pour différentes

populations de *T. brasiliensis*. Nous pourrions cibler les gènes chimiosensoriels révélés DE dans notre étude et définir des amorces spécifiques afin de réaliser des qPCR entre les différentes populations évoluant dans des conditions similaires d'élevage. L'absence de différences d'expression de ces gènes signifierait une adaptation plastique des différentes populations à ce nouvel environnement. En revanche, si des différences d'expression sont observées après plusieurs générations d'élevage entre populations issues d'individus domiciliés et populations issues d'individus sylvatiques dans le même sens que celles observées entre les populations prélevées directement sur le terrain, alors l'hypothèse de l'adaptation génétique ou d'héritabilité de marques épigénétiques serait la plus probable.

Une des perspectives qui pourrait être facilement mise en place avec le jeu de données transcriptomiques des différentes espèces de *Rhodnius* serait d'étudier la variation de pressions sélectives et la sélection positive au niveau des séquences codantes de nos gènes d'intérêt. Nous pourrions utiliser une méthode basée sur la fréquence des substitutions par le biais du programme codeml du package PAML (Yang, 2007). Ce programme estime par une méthode de maximum de vraisemblance le ratio du taux de substitution non synonyme normalisé (dN) sur le taux de substitution synonyme normalisé (dS). Un ratio significativement supérieur à 1 est considéré comme une preuve de sélection positive pour un groupe d'acides aminés, tandis qu'un ratio significativement inférieur à 1 indique de la sélection purificatrice. Nous pourrions alors estimer si certains allèles subissent des pressions de sélection en lien avec la capacité des *Rhodnius* à se domicilier. Par ailleurs, des séquençages via la méthode PacBio vont être prochainement réalisés sur les génomes des différentes espèces de *Rhodnius*. Ce plus grand jeu de données génomiques nous permettra entre autres de balayer les traces de sélection plus largement et de ne pas se restreindre à l'analyse des quelques gènes d'intérêt.

Etudes fonctionnelles des gènes chimiosensoriels

Afin d'évaluer la fonction olfactive des gènes chimiosensoriels annotés dans nos modèles de Triatominae, nous pourrions réaliser des expériences de silencing (suppression de l'expression du gène) par l'utilisation de siRNA. L'analyse des conséquences du silencing sur le comportement de punaises hématophages en réponse à l'émission d'odeurs pourrait nous permettre d'associer les gènes annotés à la reconnaissance de molécules spécifiques. Plusieurs molécules pourraient être testées : les molécules émises par différents types d'hôtes et connues pour être détectées par les punaises (CO₂, acide isobutyrique, 1-octène-3-ol et amoniac ou nonanal) ; des phéromones sexuelles, des phéromones d'alarme ou de signaux d'agrégation. Il serait alors possible d'associer les gènes chimiosensoriels aux molécules olfactives qu'ils détectent et donc d'étudier leur implication dans la recherche alimentaire, la recherche de partenaire, l'agrégation ou la fuite d'un danger.

Utilisation des gènes chimiosensoriels dans la lutte vectorielle

A l'heure actuelle, la principale stratégie de prévention contre la maladie de Chagas consiste à éradiquer chimiquement les vecteurs qui ont colonisé les habitations humaines. Cependant, des cas de résistance aux insecticides ont déjà été démontrés chez *Triatoma infestans* (González Audino et al., 2004; Mougabure-Cueto and Picollo, 2015; Orihuela et al., 2008). Il est donc impératif de développer d'autres stratégies de lutte anti-vectorielle afin de répondre aux problématiques de recolonisation post-traitement et d'anticiper l'éventuelle résistance aux insecticides que pourraient développer les vecteurs.

Nous avons pu mettre en évidence dans notre étude un profil d'expression divergent des gènes chimiosensoriels entre les espèces ou populations domiciliées et sylvatiques. Le répertoire des gènes codant les OBPs semble être varié chez les espèces domiciliées alors que l'expression de certains d'entre eux est diminuée. Or, les OBPs sont les premières protéines à intervenir dans la détection des molécules olfactives. Leur contribution dans le comportement spécifique des insectes en réponse à des signaux chimiques volatiles est largement admise. Les divergences observées entre les individus sylvatiques et domiciliés révèlent donc des modifications dans la détection des signaux olfactifs au sein des milieux anthropisés. Cibler les OBPs exprimées chez les punaises domiciliées et non détectées chez les sylvatiques afin d'interrompre précocement la perception du signal chimique, en empêchant la fixation et le transport de la molécule est une stratégie plausible pour bloquer l'accouplement des insectes et la détection des hôtes. Hooper et al., (2009) ont montré chez *Bombyx mori* qu'un composé analogue synthétique pouvait se lier à la PBP1 (protéine de liaison à une phéromone) avec une affinité supérieure à celle de la phéromone naturellement captée par cette protéine. Par ailleurs, une autre étude a démontré chez *Bombyx mori* que certaines OBPs pouvaient lier plusieurs phéromones et d'autres pouvaient être très spécifiques. Ainsi, BmorGOBP2 est capable de discriminer deux phéromones sexuelles similaires d'après leur structure à savoir le bombykol et le bombykal (Zhou et al., 2009). Cibler une OBP qui lie de manière spécifique une phéromone en particulier donne la possibilité de cibler une espèce précise, en évitant les effets sur d'autres insectes (Zhou et al., 2010b). L'utilisation de contrôles basés sur des pièges à phéromones ou de désorientation sexuelle par le biais de phéromones synthétiques a déjà fait ses preuves. Ils ont été mis en place pour de nombreuses espèces d'insectes ravageurs de cultures, et ce pour des ordres très variés tels que les coléoptères, les diptères ou les lépidoptères (Witzgall et al., 2010). Des pièges imitant la présence d'hôtes par le biais d'une combinaison de CO₂, de chaleur et de kairomone synthétique sont utilisés comme technique d'échantillonnage de punaises de lit *Cimex lectularius* et pourraient à l'avenir être améliorés en vue de lutter contre ces hémiptères (Schaafsma et al., 2012; Weeks et al., 2011). Plusieurs études montrent que l'utilisation à long terme de contrôles basés sur les phéromones aboutit à la diminution de la population ciblée (Witzgall et al., 2010). Ainsi, après avoir réalisé une étude fonctionnelle des OBPs chez les Triatominae comme suggéré ci-dessus, plusieurs stratégies de lutte contre les vecteurs de la maladie de Chagas pourraient être envisagées. On peut envisager, en périphérie des habitations,

l'utilisation de pièges comportant un leurre constitué d'analogues de phéromones ou de molécules attractives pour les punaises. Des phéromones sexuelles synthétiques engendrant une désorientation sexuelle des punaises pourraient être déposées autour et dans les habitations afin d'empêcher leur accouplement et leur prolifération autour de ces zones. Cependant, les phéromones sexuelles étant spécifiques de l'espèce, cette technologie nécessiterait l'établissement de pièges différents en fonction des espèces de Triatominae présentes dans les zones géographiques traitées.

Communications

Publications

Marchant, A., Mougél, F., Almeida, C., Jacquin-Joly, E., Costa, J., Harry, M., 2015. **Chemosensory genes are mainly under expressed in domiciliary bugs, *Triatoma brasiliensis*, compared to sylvatic and peridomiciliary populations** Submitted to PLOS Neglected Tropical Diseases

Marchant, A., Mougél, F., Mendonça, V., Quartier, M., Jacquin-Joly, E., da Rosa, J.A., Petit, E., and Harry, M., 2015. **Comparing *de novo* and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug *Rhodnius prolixus***. Insect Biochem. Mol. Biol.

Marchant, A., Mougél, F., Almeida, C., Jacquin-Joly, E., Costa, J., Harry, M., 2014. ***De novo* transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease**. Genetica 1–15. doi:10.1007/s10709-014-9790-5

Présentations Orales

British Ecological Society and Société Française d'Ecologie Annual Meeting, Lille, France- 12/2014
“**Comparative transcriptomics in Chagas disease vectors: a focus to chemosensory genes.**”
A. Marchant, F. Mougél, C. Almeida, E. Jacquin-Joly, J. Costa, M. Harry.

36ème réunion du Groupe d'Etude de Biologie et Génétique des Populations, Orsay, France- 08/2014.
“**Study of the chemosensory transcriptome of bloodsucking bugs, vectors of Chagas disease.**”
A. Marchant, F. Mougél, C. Almeida, J. Costa, E. Jacquin-Joly, M. Harry.

Bioinformatics for Environmental Genomics, Lyon, France – 05/ 2014
“***De novo* transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease.** “ A. Marchant, F. Mougél, C. Almeida, E. Jacquin-Joly, J. Costa, M. Harry

35ème réunion du Groupe d'Etude de Biologie et Génétique des Populations, Marseille, France- 08/2013
“**Chemosensory transcriptomes (OBP, CSP) of two bloodsucking bugs, Chagas disease vectors.**”
A. Marchant, F. Mougél, C. Almeida, E. Jacquin-Joly, J. Costa, M. Harry

Présentations Poster

SMBE Satellite meeting SMBEBA: Investigating biological adaptation with NGS, Montpellier, France-05/2015
“**Differential expression analysis of the chemosensory transcriptome of Chagas disease vectors**”
A. Marchant, F. Mougél, C. Almeida, E. Jacquin-Joly, J. Costa, M. Harry.

Seventh International Symposium on Molecular Insect Science, Amsterdam- 07/2014
“**Chemosensory transcriptomes of bloodsucking bugs, Chagas disease vectors**”.
A. Marchant, F. Mougél, C. Almeida, J. Costa, E. Jacquin-Joly, M. Harry

Colloque de Génomique Environnementale, Rennes- 11/2013
“**Chemosensory transcriptomes (OBP, CSP) of two bloodsucking bugs vectors of Chagas disease**”
A. Marchant, F. Mougél, C. Almeida, E. Jacquin-Joly, J. Costa, M. Harry

Annexes

Principales technologies de séquençage à haut débit

454

Le 454 est la première technologie de deuxième génération à avoir vu le jour. Cette méthode a été développée en 2005 par Johnathan M. Rothberg (Margulies et al., 2005). Elle est fondée sur l'intégration de plusieurs techniques : le pyroséquençage, les technologies des plaques en fibre optique picotitrée, la PCR en émulsion et les technologies informatiques de pointe permettant l'acquisition, le traitement et l'analyse des images. L'ADN est fragmenté et les fragments générés sont liés à des adaptateurs spécifiques. Des microbilles sont revêtues d'amorces complémentaires de ces adaptateurs ce qui permet la fixation d'un fragment d'ADN simple brin par microbille. Les microbilles contenant les fragments d'ADN sont ensuite mises en émulsion en présence des réactifs de PCR. Chaque goutte est considérée comme un microréacteur et englobe une seule microbille, donc un seul fragment d'ADN. Par conséquent, chaque fragment est amplifié séparément. Cette amplification est nécessaire pour obtenir une intensité suffisante du signal lumineux pour une détection fiable. Ensuite, chaque bille contenant un fragment d'ADN amplifié est placée sur une puce de fibres optiques sensibles. Le séquençage se déroule selon la technique du pyroséquençage. C'est-à-dire que la séquence est révélée en temps réel en lien avec l'activité de l'ADN polymérase. La lumière émise lors de la libération de pyrophosphate inorganique après incorporation d'un dNTP au brin d'ADN néosynthétisé est mesurée par le séquenceur et permet la détermination du dNTP incorporé. Ainsi, l'ordre et l'intensité des pics de lumière enregistrés comme diagrammes de barre révèlent les séquences d'ADN sous-jacentes (Mardis, 2008; Metzker, 2010). Les 400 000 réactions de séquençages peuvent être réalisées simultanément avec les techniques 454, soit approximativement 100 Mb de données de séquençage fournies en 7 heures de run (Mardis, 2008). La longueur des séquences peut aller jusqu'à 600pb. Cependant, le taux d'erreur de séquençage est supérieur à 1% dans les régions homopolymères (Ansorge, 2009).

Figure 9 : Les différentes étapes de la technologie 454

1) Construction de la librairie par fragmentation et fixation d'adaptateurs aux fragments d'ADN. 2) Amplification par PCR des fragments séparément dans des bulles d'émulsion. 3) Les billes sont chargées dans les plaques en fibre optique picotitrée. 4) Réaction de pyroséquençage basée sur la détection de signaux lumineux libérés lors de l'incorporation de nucléotide (adaptée de <http://www.454.com>).

(cf Figure 1 <http://bitesizebio.com/13548/sequencing-genomes-from-neanderthals-to-james-watson-the-roche-454-genome-sequencer-explained/>)

Illumina

C'est en 2006 que cette technologie, basée sur les inventions de S. Balasubramanian et D. Klenerman de l'Université de Cambridge a vu le jour. D'abord commercialisée par Solexa, elle est rachetée plus tard par Illumina. Cette technologie est basée sur la méthode de Sanger qui utilise des nucléotides marqués et qui bloquent l'élongation. L'ADN est tout d'abord fragmenté, puis des adaptateurs sont fixés sur chaque extrémité des fragments. Les fragments sont ensuite distribués au hasard sur la surface d'une cellule de flux déjà recouverte de petits fragments d'ADN complémentaires aux séquences des adaptateurs. Les deux extrémités des fragments d'ADN se fixent par complémentarité de séquences des adaptateurs sur la cellule formant ainsi des ponts d'amplification qui facilitent et augmentent le débit du séquençage. L'amplification du fragment est alors réalisée par PCR à la surface de cette cellule de flux en formant des groupes de fragments identiques appelés clusters. Le séquençage des fragments amplifiés est basé sur un processus cyclique qui comprend l'incorporation de nucléotides fluorescents, l'imagerie de fluorescence et le clivage. Dans une première étape, l'ADN polymérase incorpore un nucléotide modifié et marqué par fluorescent de façon complémentaire au fragment matrice. La modification « terminator » de ce nucléotide bloque la synthèse de l'ADN de façon à ce qu'une seule base soit hybridée par cycle. L'hybridation d'un nucléotide modifié et marqué est suivie par le lavage des nucléotides non utilisés et par l'acquisition d'image permettant d'identifier la nature du nucléotide incorporé. Puis, le groupement « terminator » et le fluorescent de ce nucléotide incorporé sont clivés chimiquement. L'amorce peut alors initier un nouveau cycle en incorporant le nucléotide complémentaire suivant du brin matrice. Plusieurs cycles sont alors enchaînés ainsi de suite afin de séquencer le fragment (Ansorge, 2009; Mardis, 2008). Le nombre de cycles est actuellement limité à 150 pour la technologie GenomeAnalyzer IIx et 100 pour les instruments HiSeq2000 (Mardis, 2008). Le nombre maximal de cycles est limité par des imperfections dans la chimie provoquant une incertitude accrue dans les mesures au fur et à mesure que le nombre de cycles augmente. Il en résulte une réduction de la qualité des bases au fur et à mesure des cycles. La très haute densité de la puce (plus de 100 millions de molécules par centimètre carré) permet de séquencer une quantité très importante de bases (environ 100 000) par secondes. Il est également possible de lire les fragments à partir des deux extrémités (séquençage paired-end) ; la distance entre les paires de reads est alors déterminée par la distribution de la taille des fragments. Ce type de séquençage fonctionne avec des fragments de tailles comprises entre 100-500bp (SNIC-UPPMAX, 2013). Pour de plus longues tailles d'inserts (2-5kb) une technique appelée « séquençage Mate-Paire » peut être appliquée.

Un des avantages de cette technologie par rapport à celle du 454 réside dans le nombre très important de lectures qu'elle génère. Par conséquent, la profondeur de séquençage, c'est-à-dire le nombre de fois qu'un site est séquençé, est bien plus importante que celle fournie par le

séquençage 454. De plus, le 454 génère au niveau des homopolymères des erreurs de séquençage que ne rencontre pas l'Illumina. En revanche, la longueur des reads générée par l'Illumina (en moyenne 100bp) reste réduite par rapport aux reads de 454 (500 à 600pb).

Figure 10 : Les différentes étapes de l'Illumina.

Après fragmentation de l'ADN et fixation des adaptateurs, la librairie est ajoutée à une cellule de flux pour le procéder à la mise en place du pont d'amplification. Les fragments sont ensuite dénaturés, hybridés avec une amorce de séquençage et soumis à un séquençage par synthèse en utilisant des nucléotides bloqués en 3' et marqués par fluorochromes.(cf <https://www.uppmass.uu.se/illumina-sequencing>)

Le séquençage par ligature et détection Oligo (SOLID)

En 2007, Applied Biosystems (aujourd'hui repris par Life Technologies) propose le séquençage SOLID. Ce processus démarre un peu comme pour le pyroséquençage du 454, à savoir qu'une librairie de brins d'ADN est associée à une microbille magnétique d'un micron d'épaisseur *via* un adaptateur. Chaque brin est amplifié par PCR à émulsion. Les brins d'ADN amplifiés sont ensuite fixés sur une plaque de verre et chaque fragment est associé à une amorce pour pouvoir initialiser sa réplication. La réplication des fragments d'ADN s'appuie sur la ligation, un processus habituellement utilisé pour la réparation de l'ADN des cellules, notamment lors de la réplication naturelle. La ligase est une protéine complexe capable de raccommoier des morceaux d'ADN les uns aux autres. La technique consiste à hybrider un bloc de 8 paires (8mer) de bases comprenant des bases marquées par fluorescence ainsi que des bases ATGC non marquée. Après l'étape de ligation, un affichage fluorescent identifie les bases marquées. Puis une étape de clivage chimique supprime les bases universelles et l'opération est répétée en décalant le tout d'une base. Cela permet par recoupement d'identifier les bases de l'ADN et de les mesurer chacune deux fois. Ceci réduit le taux d'erreur de séquençage à 1 pour 1000, soit dix fois mieux que dans le pyroséquençage. Ce type de séquençage nécessite environ 5 jours de run et produit entre 3 et 4 Go de données de séquences avec une longueur moyenne de lecture de 25 à 35 pb (Mardis, 2008; Metzker, 2010).

Figure 11 : Fonctionnement du séquençage SOLID (adaptée de Metzker, 2010) (cf Figure 3 Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1), 31-46.)

La 3^{ème} génération de séquençage

La technologie Pacific Biosciences est actuellement le leader de la 3^{ème} génération de séquençage. Elle permet de séquencer des fragments d'ADN de très grande taille (jusqu'à 2500 nucléotides) à partir d'une seule molécule d'ADN. Un catalyseur biologique permettant l'élongation de l'ADN est fixé au fond d'un puits. Il va incorporer un par un des nucléotides libres complémentaires du brin à analyser. Les nucléotides mis à disposition du catalyseur sont marqués par des molécules fluorescentes de différentes couleurs (bleu pour G, jaune pour A, vert pour T et rouge pour C). Lors de l'incorporation d'un nouveau nucléotide par le

catalyseur, le marqueur fluorescent va être libéré et générer un signal lumineux. Ce signal sera détecté par le séquenceur dans une zone restreinte au fond du puits. On obtient alors une multitude de points colorés, correspondants à l'alignement des puits sur le support et aux nucléotides incorporés dans chacun de ces puits. Les images obtenues seront enregistrées et analysées par des méthodes informatiques de pointe (Metzker, 2010).

Figure 12 : Séquençage de troisième génération : Pacific Biosciences (adaptée de Metzker, 2010) (cf Figure 4 Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1),

Références

- Abad-Franch, F., and Monteiro, F.A. (2007). Biogeography and evolution of Amazonian triatomines (Heteroptera: Reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. *Mem. Inst. Oswaldo Cruz* 102, 57–70.
- Abad-Franch, F., Monteiro, F.A., Jaramillo O., N., Gurgel-Gonçalves, R., Dias, F.B.S., and Diotaiuti, L. (2009). Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: A multi-scale appraisal of the tribe Rhodniini (Triatominae). *Acta Trop.* 110, 159–177.
- Abad-Franch, F., Ferraz, G., Campos, C., Palomeque, F.S., Grijalva, M.J., Aguilar, H.M., and Miles, M.A. (2010). Modeling Disease Vector Occurrence when Detection Is Imperfect: Infestation of Amazonian Palm Trees by Triatomine Bugs at Three Spatial Scales. *PLoS Negl Trop Dis* 4, e620.
- Abad-Franch, F., Pavan, M.G., Jaramillo-O, N., Palomeque, F.S., Dale, C., Chaverra, D., Monteiro, F.A., Abad-Franch, F., Pavan, M.G., Jaramillo-O, N., et al. (2013). *Rhodnius barretti*, a new species of Triatominae (Hemiptera: Reduviidae) from western Amazonia. *Mem. Inst. Oswaldo Cruz* 108, 92–99.
- Abraham, L.B., Gorla, D.E., and Catalá, S.S. (2011). Dispersal of *Triatoma infestans* and other Triatominae species in the arid Chaco of Argentina: Flying, walking or passive carriage? The importance of walking females. *Mem. Inst. Oswaldo Cruz* 106, 232–239.
- Abuin, L., Bargeton, B., Ulbrich, M.H., Isacoff, E.Y., Kellenberger, S., and Benton, R. (2011). Functional Architecture of Olfactory Ionotropic Glutamate Receptors. *Neuron* 69, 44–60.
- Acín, P., Carrascal, M., Abián, J., Guerrero, A., and Quero, C. (2009). Expression of differential antennal proteins in males and females of an important crop pest, *Sesamia nonagrioides*. *Insect Biochem. Mol. Biol.* 39, 11–19.
- Agrawal, A.A. (2001). Phenotypic Plasticity in the Interactions and Evolution of Species. *Science* 294, 321–326.
- Aguilar, H.M., Abad-Franch, F., Dias, J.C.P., Junqueira, A.C.V., and Coura, J.R. (2007). Chagas disease in the Amazon Region. *Mem. Inst. Oswaldo Cruz* 102, 47–56.
- Almeida, C.E., Faucher, L., Lavina, M., Costa, J., and Harry, M. (submitted). Molecular eco-epidemiology of *Triatoma brasiliensis* using an individually based approach: Inferences on Chagas disease transmission links. *Mol. Ecol.*
- Almeida, C.E., Duarte, R., Nascimento, R.G. do, Pacheco, R.S., and Costa, J. (2002). *Triatoma rubrovaria* (Blanchard, 1843) (Hemiptera, Reduviidae, Triatominae) II: trophic resources and ecological observations of five populations collected in the State of Rio Grande do Sul, Brazil. *Mem. Inst. Oswaldo Cruz* 97, 1127–1131.
- Almeida, C.E., Pacheco, R.S., Haag, K., Dupas, S., Dotson, E.M., and Costa, J. (2008). Inferring from the Cyt B Gene the *Triatoma brasiliensis* Neiva, 1911 (Hemiptera: Reduviidae: Triatominae) Genetic Structure and Domiciliary Infestation in the State of Paraíba, Brazil. *Am. J. Trop. Med. Hyg.* 78, 791–802.
- Almeida, C.E., Folly-Ramos, E., Peterson, A.T., Lima-Neiva, V., Gumiel, M., Duarte, R., Lima, M.M., Locks, M., Beltrão, M., and Costa, J. (2009). Could the bug *Triatoma sherlocki* be vectoring Chagas disease in small mining communities in Bahia, Brazil? *Med. Vet. Entomol.* 23, 410–417.
- Altincicek, B., and Vilcinskis, A. (2007). Identification of immune-related genes from an apterygote insect, the firebrat *Thermobia domestica*. *Insect Biochem. Mol. Biol.* 37, 726–731.
- Andersen, S.O. (1998). Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* 28, 421–434.
- Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R., and Pelosi, P. (1999). Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur. J. Biochem.* 262, 745–754.
- Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. *New Biotechnol.* 25, 195–203.
- Assumpção, T.C.F., Francischetti, I.M.B., Andersen, J.F., Schwarz, A., Santana, J.M., and Ribeiro, J.M.C. (2008). An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas’ disease. *Insect Biochem. Mol. Biol.* 38, 213–232.
- Assumpção, T.C.F., Charneau, S., Santiago, P.B.M., Francischetti, I.M.B., Meng, Z., Araújo, C.N., Pham, V.M., Queiroz, R.M.L., Nery de Castro, C., André Ricart, C., et al. (2011). Insight into the Salivary Transcriptome and Proteome of *Dipetalogaster maxima*. *J. Proteome Res.* 10, 669–679.
- Aznar, C., La Ruche, G., Laventure, S., Carme, B., Liegeard, P., and Hontebeyrie, M. (2004). Seroprevalence of *Trypanosoma cruzi* infection in French Guiana. *Mem. Inst. Oswaldo Cruz* 99, 805–808.

- Bai, X., Mamidala, P., Rajarapu, S.P., Jones, S.C., and Mittapalli, O. (2011). Transcriptomics of the Bed Bug (*Cimex lectularius*). *PLoS ONE* *6*, e16336.
- Ban, L., Scaloni, A., Brandazza, A., Angeli, S., Zhang, L., Yan, Y., and Pelosi, P. (2003). Chemosensory proteins of *Locusta migratoria*. *Insect Mol. Biol.* *12*, 125–134.
- Barbu, C., Dumonteil, E., and Gourbière, S. (2009). Optimization of Control Strategies for Non-Domiciliated *Triatoma dimidiata*, Chagas Disease Vector in the Yucatán Peninsula, Mexico. *PLoS Negl Trop Dis* *3*, e416.
- Barghini, A., and de Medeiros, B.A.S. (2010). Artificial Lighting as a Vector Attractant and Cause of Disease Diffusion. *Environ. Health Perspect.*
- Bargues, M.D., Klisiowicz, D.R., Panzera, F., Noireau, F., Marcilla, A., Perez, R., Rojas, M.G., O'Connor, J.E., Gonzalez-Candelas, F., Galvão, C., et al. (2006). Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect. Genet. Evol.* *6*, 46–62.
- Barreto-de-Albuquerque, J., Silva-dos-Santos, D., Pérez, A.R., Berbert, L.R., Santana-van-Vliet, E. de, Farias-de-Oliveira, D.A., Moreira, O.C., Roggero, E., Carvalho-Pinto, C.E. de, Jurberg, J., et al. (2015). *Trypanosoma cruzi* Infection through the Oral Route Promotes a Severe Infection in Mice: New Disease Form from an Old Infection? *PLoS Negl Trop Dis* *9*, e0003849.
- Barrozo, R.B., and Lazzari, C.R. (2004a). The response of the blood-sucking bug *Triatoma infestans* to carbon dioxide and other host odours. *Chem. Senses* *29*, 319–329.
- Barrozo, R.B., and Lazzari, C.R. (2004b). Orientation behaviour of the blood-sucking bug *Triatoma infestans* to short-chain fatty acids: Synergistic effect of L-lactic acid and carbon dioxide. *Chem. Senses* *29*, 833–841.
- Barrozo, R.B., Minoli, S.A., and Lazzari, C.R. (2004). Circadian rhythm of behavioural responsiveness to carbon dioxide in the blood-sucking bug *Triatoma infestans* (Heteroptera : Reduviidae). *J. Insect Physiol.* *50*, 249–254.
- Batista, V.S.P., Fernandes, F.A., Cordeiro-Estrela, P., Sarquis, O., and Lima, M.M. (2013). Ecotope effect in *Triatoma brasiliensis* (Hemiptera: Reduviidae) suggests phenotypic plasticity rather than adaptation. *Med. Vet. Entomol.* *27*, 247–254.
- Benchimol Barbosa, P.R. (2006). The oral transmission of Chagas' disease: An acute form of infection responsible for regional outbreaks. *Int. J. Cardiol.* *112*, 132–133.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* *57*, 289–300.
- Benoit, J.B., Lopez-Martinez, G., Phillips, Z.P., Patrick, K.R., and Denlinger, D.L. (2010). Heat shock proteins contribute to mosquito dehydration tolerance. *J. Insect Physiol.* *56*, 151–156.
- Benton, R. (2015). Multigene Family Evolution: Perspectives from Insect Chemoreceptors. *Trends Ecol. Evol.* *30*, 590–600.
- Benton, R., Sachse, S., Michnick, S.W., and Vossall, L.B. (2006). Atypical Membrane Topology and Heteromeric Function of *Drosophila* Odorant Receptors In Vivo. *PLoS Biol* *4*, e20.
- Benton, R., Vannice, K.S., Gomez-Diaz, C., and Vossall, L.B. (2009). Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. *Cell* *136*, 149–162.
- Béranger, J.-M., and Pluot-Sigwalt, D. (2002). *Rhodnius amazonicus* Almeida, Santos & Sposina, 1973, Bona Species, Close to *R. pictipes* Stål, 1872 (Heteroptera, Reduviidae, Triatominae). *Mem. Inst. Oswaldo Cruz* *97*, 73–77.
- Béranger, J.-M., Pluot-Sigwalt, D., Pagès, F., Blanchet, D., and Aznar, C. (2009). The triatominae species of French Guiana (Heteroptera: Reduviidae). *Mem. Inst. Oswaldo Cruz* *104*, 1111–1116.
- Bergé, J., Feyereisen, R., and Amichot, M. (1998). Cytochrome P450 monooxygenases and insecticide resistance in insects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *353*, 1701–1705.
- Blomberg, S.P., and Garland, T. (2002). Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *J. Evol. Biol.* *15*, 899–910.
- Bodin, A., Barrozo, R.B., Couton, L., and Lazzari, C.R. (2008). Temporal modulation and adaptive control of the behavioural response to odours in *Rhodnius prolixus*. *J. Insect Physiol.* *54*, 1343–1348.
- Bodin, A., Vinauger, C., and Lazzari, C.R. (2009a). State-dependency of host-seeking in *Rhodnius prolixus*: The post-ecdysis time. *J. Insect Physiol.* *55*, 574–579.
- Bodin, A., Vinauger, C., and Lazzari, C.R. (2009b). Behavioural and physiological state dependency of host seeking in the blood-sucking insect *Rhodnius prolixus*. *J. Exp. Biol.* *212*, 2386–2393.
- Bohbot, J.D., and Pitts, R.J. (2015). The narrowing olfactory landscape of insect odorant receptors. *Chem. Ecol.* *3*, 39.

- Bohbot, J., Pitts, R.J., Kwon, H.-W., Rützler, M., Robertson, H.M., and Zwiebel, L.J. (2007). Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol. Biol.* *16*, 525–537.
- Bonen, L. (1993). Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* *7*, 40–46.
- Boonsuepsakul, S., Luepromchai, E., and Rongnoparut, P. (2008). Characterization of *Anopheles minimus* CYP6AA3 expressed in a recombinant baculovirus system. *Arch. Insect Biochem. Physiol.* *69*, 13–21.
- Borges, É.C., Dujardin, J.-P., Schofield, C.J., Romanha, A.J., and Diotaiuti, L. (2005). Dynamics between sylvatic, peridomestic and domestic populations of *Triatoma brasiliensis* (Hemiptera: Reduviidae) in Ceará State, Northeastern Brazil. *Acta Trop.* *93*, 119–126.
- Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., et al. (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* *2*, 1–31.
- Brand, P., Ramírez, S.R., Leese, F., Quezada-Euan, J.J., Tollrian, R., and Eltz, T. (2015). Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC Evol. Biol.* *15*, 176.
- Bray, S., and Amrein, H. (2003). A Putative *Drosophila* Pheromone Receptor Expressed in Male-Specific Taste Neurons Is Required for Efficient Courtship. *Neuron* *39*, 1019–1029.
- Breniere, S.F., Bosseno, M.-F., Vargas, F., Yaksic, N., Noireau, F., Noel, S., Dujardin, J.-P., and Tibayrenc, M. (1998). Smallness of the Panmictic Unit of *Triatoma infestans* (Hemiptera: Reduviidae). *J. Med. Entomol.* *35*, 911–917.
- Buchan, D.W.A., Ward, S.M., Lobley, A.E., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2010). Protein annotation and modelling servers at University College London. *Nucleic Acids Res.* *38*, W563–W568.
- Burgess, D.J. (2013). Genetic screens: RNA-seq into the toolkit. *Nat. Rev. Genet.* *14*, 154–155.
- Burton, V., Mitchell, H.K., Young, P., and Petersen, N.S. (1988). Heat shock protection against cold stress of *Drosophila melanogaster*. *Mol. Cell. Biol.* *8*, 3550–3552.
- Bussacos, A.C.M., Nakayasu, E.S., Hecht, M.M., Parente, J.A., Soares, C.M.A., Teixeira, A.R.L., and Almeida, I.C. (2011). Diversity of anti-haemostatic proteins in the salivary glands of *Rhodnius* species transmitters of Chagas disease in the greater Amazon. *J. Proteomics* *74*, 1664–1672.
- Cabrera, R., Vega, S., Valderrama, Y., Cabanillas, K., Fernandez, C., Rodriguez, O., Del Aguila, C., Hernandez, J., Mendoza, L., Ramon Meza, J., et al. (2013). New focus of active transmission of Chagas disease in indigenous populations in the Peruvian Amazon basin. *Rev. Soc. Bras. Med. Trop.* *46*, 367–372.
- Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., and Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Resour.* *12*, 834–845.
- Calatayud, P.-A., Marion-Poll, F., and Thiéry, D. (2013). La réception sensorielle chez les insectes. In *Interactions insectes-plantes*, (Editions Quae),.
- Calvo, E., Mans, B.J., Ribeiro, J.M.C., and Andersen, J.F. (2009). Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 3728–3733.
- Campanacci, V., Mosbah, A., Bornet, O., Wechselberger, R., Jacquin-Joly, E., Cambillau, C., Darbon, H., and Tegoni, M. (2001). Chemosensory protein from the moth *Mamestra brassicae*. *Eur. J. Biochem.* *268*, 4731–4739.
- Canini, L. (2010). Les zoonoses en France : évaluation des connaissances des médecins et vétérinaires. other.
- Carbajal de la Fuente, A.L., Minoli, S.A., Lopes, C.M., Noireau, F., Lazzari, C.R., and Lorenzo, M.G. (2007). Flight dispersal of the Chagas disease vectors *Triatoma brasiliensis* and *Triatoma pseudomaculata* in northeastern Brazil. *Acta Trop.* *101*, 115–119.
- Cariño, F.A., Koener, J.F., Plapp Jr., F.W., and Feyereisen, R. (1994). Constitutive overexpression of the cytochrome P450 gene CYP6A1 in a house fly strain with metabolic resistance to insecticides. *Insect Biochem. Mol. Biol.* *24*, 411–418.
- Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and McQuillan, J.A. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* *28*, 464–469.
- Castro, M.C., Barrett, T.V., Santos, W.S., Abad-Franch, F., and Rafael, J.A. (2010). Attraction of Chagas disease vectors (Triatominae) to artificial light sources in the canopy of primary Amazon rainforest. *Mem. Inst. Oswaldo Cruz* *105*, 1061–1064.

- Chevreur, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller, W.E.G., Wetter, T., and Suhai, S. (2004). Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Res.* 14, 1147–1159.
- Ching, T., Huang, S., and Garmire, L.X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* 20, 1684–1696.
- Chyb, S., Dahanukar, A., Wickens, A., and Carlson, J.R. (2003). Drosophila Gr5a encodes a taste receptor tuned to trehalose. *Proc. Natl. Acad. Sci.* 100, 14526–14530.
- Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450, 203–218.
- Clark, S., Egan, R., Frazier, P.I., and Wang, Z. (2013). ALE: a Generic Assembly Likelihood Evaluation Framework for Assessing the Accuracy of Genome and Metagenome Assemblies. *Bioinformatics* bts723.
- Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., and Carlson, J.R. (1999). A Novel Family of Divergent Seven-Transmembrane Proteins: Candidate Odorant Receptors in Drosophila. *Neuron* 22, 327–338.
- Clyne, P.J., Warr, C.G., and Carlson, J.R. (2000). Candidate Taste Receptors in Drosophila. *Science* 287, 1830–1834.
- Conesa (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21.
- Cornette, R., Kanamori, Y., Watanabe, M., Nakahara, Y., Gusev, O., Mitsumasu, K., Kadono-Okuda, K., Shimomura, M., Mita, K., Kikawada, T., et al. (2010). Identification of Anhydrobiosis-related Genes from an Expressed Sequence Tag Database in the Cryptobiotic Midge *Polypedilum vanderplanki* (Diptera; Chironomidae). *J. Biol. Chem.* 285, 35889–35899.
- Costa, J. (1999). The synanthropic process of Chagas disease vectors in Brazil, with special attention to *Triatoma brasiliensis* Neiva, 1911 (Hemiptera, Reduviidae, Triatominae) population, genetical, ecological, and epidemiological aspects. *Mem. Inst. Oswaldo Cruz* 94, 239–241.
- Costa, J., and Lorenzo, M. (2009). Biology, diversity and strategies for the monitoring and control of triatomines - Chagas disease vectors. *Mem. Inst. Oswaldo Cruz* 104, 46–51.
- Costa, J., Barth, O.M., Marchon-Silva, V., de Almeida, C.E., Freitas-Sibajev, M.G.R., and Panzera, F. (1997). Morphological Studies on the *Triatoma brasiliensis* Neiva, 1911 (Hemiptera, Reduviidae, Triatominae) Genital Structures and Eggs of Different Chromatic Forms. *Mem. Inst. Oswaldo Cruz* 92, 493–498.
- Costa, J., de Almeida, J.R., Britto, C., Duarte, R., Marchon-Silva, V., and Pacheco, R. da S. (1998). Ecotopes, natural infection and trophic resources of *Triatoma brasiliensis* (Hemiptera, Reduviidae, Triatominae). *Mem. Inst. Oswaldo Cruz* 93, 7–13.
- Costa, J., Almeida, C.E., Dotson, E.M., Lins, A., Vinhaes, M., Silveira, A.C., and Beard, C.B. (2003). The epidemiologic importance of *Triatoma brasiliensis* as a Chagas disease vector in Brazil: a revision of domiciliary captures during 1993-1999. *Mem. Inst. Oswaldo Cruz* 98, 443–449.
- Costa, J., Peterson, A.T., and Dujardin, J.P. (2009). Morphological evidence suggests homoploid hybridization as a possible mode of speciation in the Triatominae (Hemiptera, Heteroptera, Reduviidae). *Infect. Genet. Evol.* 9, 263–270.
- Costa, J., Correia, N.C., Neiva, V.L., Gonçalves, T.C.M., and Felix, M. (2013). Revalidation and redescription of *Triatoma brasiliensis* macromelasoma Galvão, 1956 and an identification key for the *Triatoma brasiliensis* complex (Hemiptera: Reduviidae: Triatominae). *Mem. Inst. Oswaldo Cruz* 108, 785–789.
- Costa, J., Dornak, L.L., Almeida, C.E., and Peterson, A.T. (2014). Distributional potential of the *Triatoma brasiliensis* species complex at present and under scenarios of future climate conditions. *Parasit. Vectors* 7, 238.
- Costa-da-Silva, A.L., Kojin, B.B., Marinotti, O., James, A.A., and Capurro, M.L. (2013). Expression and accumulation of the two-domain odorant-binding protein AegOBP45 in the ovaries of blood-fed *Aedes aegypti*. *Parasit. Vectors* 6, 364.
- Coura, J.R. (2007). Chagas disease: what is known and what is needed - A background article. *Mem. Inst. Oswaldo Cruz* 102, 113–122.
- Coura, J.R., and Castro, S.L. de (2002). A Critical Review on Chagas Disease Chemotherapy. *Mem. Inst. Oswaldo Cruz* 97, 3–24.
- Coura, J.R., and Coura, J.R. (2013). Chagas disease: control, elimination and eradication. Is it possible? *Mem. Inst. Oswaldo Cruz* 108, 962–967.
- Coura, J.R., and Viñas, P.A. (2010). Chagas disease: a new worldwide challenge. *Nature* 465, S6–S7.

- Coura, J.R., Junqueira, A.C.V., Fernandes, O., Valente, S.A.S., and Miles, M.A. (2002). Emerging Chagas disease in Amazonian Brazil. *Trends Parasitol.* *18*, 171–176.
- Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T.J., and Benton, R. (2010). Ancient Protostome Origin of Chemosensory Ionotropic Glutamate Receptors and the Evolution of Insect Taste and Olfaction. *PLoS Genet* *6*, e1001064.
- Croucher, P.J., Brewer, M.S., Winchell, C.J., Oxford, G.S., and Gillespie, R.G. (2013). De novo characterization of the gene-rich transcriptomes of two color-polymorphic spiders, *Theridion grallator* and *T. californicum* (Araneae: Theridiidae), with special reference to pigment genes. *BMC Genomics* *14*, 862.
- Daborn, P.J., Lumb, C., Boey, A., Wong, W., French-Constant, R.H., and Batterham, P. (2007). Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression. *Insect Biochem. Mol. Biol.* *37*, 512–519.
- Dahanukar, A., Foster, K., van der Goes van Naters, W.M., and Carlson, J.R. (2001). A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. *Nat. Neurosci.* *4*, 1182–1186.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* *27*, 2156–2158.
- Dani, F.R., Michelucci, E., Francese, S., Mastrobuoni, G., Cappellozza, S., Marca, G.L., Niccolini, A., Felicioli, A., Moneti, G., and Pelosi, P. (2011). Odorant-Binding Proteins and Chemosensory Proteins in Pheromone Detection and Release in the Silkmoth *Bombyx mori*. *Chem. Senses* *36*, 335–344.
- Dansa-Petretski, M., Ribeiro, J.M.C., Atella, G.C., Masuda, H., and Oliveira, P.L. (1995). Antioxidant Role of *Rhodnius prolixus* Heme-binding Protein PROTECTION AGAINST HEME-INDUCED LIPID PEROXIDATION. *J. Biol. Chem.* *270*, 10893–10896.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* *27*, 1164–1165.
- Dauwalder, B., Tsujimoto, S., Moss, J., and Mattox, W. (2002). The *Drosophila* takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. *Genes Dev.* *16*, 2879–2892.
- David, J.-P., Coissac, E., Melodelima, C., Poupardin, R., Riaz, M.A., Chandor-Proust, A., and Reynaud, S. (2010). Transcriptome response to pollutants and insecticides in the dengue vector *Aedes aegypti* using next-generation sequencing technology. *BMC Genomics* *11*, 216.
- David, J.-P., Faucon, F., Chandor-Proust, A., Poupardin, R., Riaz, M., Bonin, A., Navratil, V., and Reynaud, S. (2014). Comparative analysis of response to selection with three insecticides in the dengue mosquito *Aedes aegypti* using mRNA sequencing. *BMC Genomics* *15*, 174.
- Dias, F.B.S., Jaramillo-O, N., Diotaiuti, L., Dias, F.B.S., Jaramillo-O, N., and Diotaiuti, L. (2014). Description and characterization of the melanic morphotype of *Rhodnius nasutus* Stål, 1859 (Hemiptera: Reduviidae: Triatominae). *Rev. Soc. Bras. Med. Trop.* *47*, 637–641.
- Dias, J.C.P., Silveira, A.C., and Schofield, C.J. (2002). The impact of Chagas disease control in Latin America: a review. *Mem. Inst. Oswaldo Cruz* *97*, 603–612.
- Díaz, S., Panzera, F., Jaramillo-O, N., Pérez, R., Fernández, R., Vallejo, G., Saldaña, A., Calzada, J.E., Triana, O., and Gómez-Palacio, A. (2014). Genetic, Cytogenetic and Morphological Trends in the Evolution of the *Rhodnius* (Triatominae: Rhodniini) Trans-Andean Group. *PLoS ONE* *9*.
- De Diego, J.A., Penin, P., Del Rey, J., Mayer, R., and Gamallo, C. (1991). A comparative pathological study of three strains of *Trypanosoma cruzi* in an experimental model.
- Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., et al. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* *14*, 671–683.
- Dopman, E.B., and Hartl, D.L. (2007). A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* *104*, 19920–19925.
- Dujardin, Schofield, and Tibayrenc (1998). Population structure of Andean *Triatoma infestans*: allozyme frequencies and their epidemiological relevance. *Med. Vet. Entomol.* *12*, 20–29.
- Dujardin, J.P., Tibayrenc, M., Venegas, E., Maldonado, L., Desjeux, P., and Ayala, F.J. (1987). Isozyme Evidence of Lack of Speciation Between Wild and Domestic *Triatoma infestans* (Heteroptera: Reduviidae) in Bolivia. *J. Med. Entomol.* *24*, 40–45.
- Dujardin, J.P., Panzera, P., and Schofield, C.J. (1999). Triatominae as a model of morphological plasticity under ecological pressure. *Mem. Inst. Oswaldo Cruz* *94*, 223–228.
- Dujardin, J.-P., Schofield, C.J., and Panzera, F. (2000). Les vecteurs de la maladie de Chagas. *Rech. Taxon. Biol. Génétiques Brux. Acad. R. Sci. Douv. Mer* *162*.

- Dumonteil, E., Ruiz-Piña, H., Rodriguez-Félix, E., Barrera-Pérez, M., Ramirez-Sierra, M.J., Rabinovich, J.E., and Menu, F. (2004). Re-infestation of houses by *Triatoma dimidiata* after intra-domicile insecticide application in the Yucatán peninsula, Mexico. *Mem. Inst. Oswaldo Cruz* *99*, 253–256.
- Dumonteil, E., Tripet, F., Ramirez-Sierra, M.J., Payet, V., Lanzaro, G., and Menu, F. (2007). Assessment of *Triatoma dimidiata* dispersal in the Yucatan Peninsula of Mexico by morphometry and microsatellite markers. *Am. J. Trop. Med. Hyg.* *76*, 930–937.
- Duvallet, G. (2006). Parasites, vecteurs de pathogènes et changements climatiques. *Hydroécologie Appliquée* *15*, 87–96.
- Duvaux, L., Geissmann, Q., Gharbi, K., Zhou, J.-J., Ferrari, J., Smadja, C.M., and Butlin, R.K. (2015). Dynamics of Copy Number Variation in Host Races of the Pea Aphid. *Mol. Biol. Evol.* *32*, 63–80.
- Eklblom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* *107*, 1–15.
- Elmore, T., Ignell, R., Carlson, J.R., and Smith, D.P. (2003). Targeted Mutation of a *Drosophila* Odor Receptor Defines Receptor Requirement in a Novel Class of Sensillum. *J. Neurosci.* *23*, 9906–9912.
- Emperaire, L., and Romana, C. (2006). Triatominae et Cactaceae : un risque pour la transmission de la Trypanosomose américaine dans le périodomicile (nord-est du Brésil). *Parasite* *13*, 171–178.
- Enayati, A.A., Ranson, H., and Hemingway, J. (2005). Insect glutathione transferases and insecticide resistance. *Insect Mol. Biol.* *14*, 3–8.
- Engsontia, P., Sanderson, A.P., Cobb, M., Walden, K.K.O., Robertson, H.M., and Brown, S. (2008). The red flour beetle's large nose: An expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* *38*, 387–397.
- Engsontia, P., Sangket, U., Chotigeat, W., and Satasook, C. (2014). Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. *J. Mol. Evol.* *79*, 21–39.
- Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* *13*, 97–109.
- Feldmeyer, B., Wheat, C.W., Krezdorn, N., Rotter, B., and Pfenninger, M. (2011). Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance. *BMC Genomics* *12*, 317.
- Feliciangeli, M.D., Dujardin, J.-P., Bastrenta, B., Mazzarri, M., Villegas, J., Flores, M., and Munoz, M. (2002). Is *Rhodnius robustus* (Hemiptera: Reduviidae) responsible for Chagas disease transmission in western Venezuela? *Trop. Med. Int. Health* *7*, 280–287.
- Feliciangeli, M.D., Campbell-Lendrum, D., Martinez, C., Gonzalez, D., Coleman, P., and Davies, C. (2003). Chagas disease control in Venezuela: lessons for the Andean region and beyond. *Trends Parasitol.* *19*, 44–49.
- Feliciangeli, M.D., Sanchez-Martin, M., Marrero, R., Davies, C., and Dujardin, J.-P. (2007). Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the State of Barinas (Venezuela). *Acta Trop.* *101*, 169–177.
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* *39*, 783–791.
- Figueiras, A.N.L., Kenigsten, A., and Lazzari, C.R. (1994). Aggregation in the haematophagous bug *Triatoma infestans*: Chemical signals and temporal pattern. *J. Insect Physiol.* *40*, 311–316.
- Finseth, F.R., and Harrison, R.G. (2014). A Comparison of Next Generation Sequencing Technologies for Transcriptome Assembly and Utility for RNA-Seq in a Non-Model Bird. *PLoS ONE* *9*, e108550.
- Fitzpatrick, S., Feliciangeli, M.D., Sanchez-Martin, M.J., Monteiro, F.A., and Miles, M.A. (2008). Molecular Genetics Reveal That Silvatic *Rhodnius prolixus* Do Colonise Rural Houses. *PLoS Negl Trop Dis* *2*, e210.
- Forêt, S., and Maleszka, R. (2006). Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res.* *16*, 1404–1413.
- Forêt, S., Wanner, K.W., and Maleszka, R. (2007). Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling. *Insect Biochem. Mol. Biol.* *37*, 19–28.
- Fox, A.N., Pitts, R.J., Robertson, H.M., Carlson, J.R., and Zwiebel, L.J. (2001). Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc. Natl. Acad. Sci.* *98*, 14693–14697.
- Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C., and Haddock, S.H.D. (2013). A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics* *14*, 167.

- Franco, T.A., Oliveira, D.S., Moreira, M.F., Leal, W.S., and Melo, A.C.A. (2015). Silencing the odorant receptor co-receptor RproOrco affects the physiology and behavior of the Chagas disease vector *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.*
- Fujii, S. (2002). Genes expressed in the *Drosophila* head reveal a role for fat cells in sex-specific physiology. *EMBO J.* *21*, 5353–5363.
- Fujii, S., Toyama, A., and Amrein, H. (2008). A Male-Specific Fatty Acid ω -Hydroxylase, SXE1, Is Necessary for Efficient Male Mating in *Drosophila melanogaster*. *Genetics* *180*, 179–190.
- Fujikawa, K., Seno, K., and Ozaki, M. (2006). A novel Takeout-like protein expressed in the taste and olfactory organs of the blowfly, *Phormia regina*. *FEBS J.* *273*, 4311–4321.
- Galindo, K., and Smith, D.P. (2001). A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* *159*, 1059–1072.
- Galvão, C., and Justi, S.A. (2015). An overview on the ecology of Triatominae (Hemiptera:Reduviidae). *Acta Trop.*
- GALVÃO, C., CARCAVALLO, R., DA SILVA ROCHA, D., and JURBERG, J. (2003). A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *ZOOTAXA* 1–36.
- Gamboa, J.C. (1973). La población silvestre de *Rhodnius prolixus* en Venezuela. *Arch Venez Med Trop Parasit Med* *5*, 321–352.
- Gao, Q., and Chess, A. (1999). Identification of Candidate *Drosophila* Olfactory Receptors from Genomic DNA Sequence. *Genomics* *60*, 31–39.
- Gardiner, A., Barker, D., Butlin, R.K., Jordan, W.C., and Ritchie, M.G. (2008a). *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol. Ecol.* *17*, 1648–1657.
- Gardiner, A., Barker, D., Butlin, R.K., Jordan, W.C., and Ritchie, M.G. (2008b). Evolution of a Complex Locus: Exon Gain, Loss and Divergence at the Gr39a Locus in *Drosophila*. *PLoS ONE* *3*, e1513.
- Gaunt, M., and Miles, M. (2000). The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Mem. Inst. Oswaldo Cruz* *95*, 557–565.
- Ghalambor, C.K., McKAY, J.K., Carroll, S.P., and Reznick, D.N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* *21*, 394–407.
- Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N., and Hughes, K.A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* *525*, 372–375.
- Gilber, S.R., Alban, S.M., Gobor, L., Bescrovaine, J. de O., Myiazaki, M.I., Thomaz-Soccol, V., Gilber, S.R., Alban, S.M., Gobor, L., Bescrovaine, J. de O., et al. (2013). Comparison of conventional serology and PCR methods for the routine diagnosis of *Trypanosoma cruzi* infection. *Rev. Soc. Bras. Med. Trop.* *46*, 310–315.
- Githeko, A.K., Lindsay, S.W., Confalonieri, U.E., and Patz, J.A. (2000). Climate change and vector-borne diseases: a regional analysis. *Bull. World Health Organ.* *78*, 1136–1147.
- Glaser, N. (2013). Contribution des sens chimiques à l'adaptation à l'environnement : une étude transcriptomique chez la noctuelle *Sesamia nonagrioides* (Lepidoptera : Noctuidae). L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech).
- Glaser, N., Gallot, A., Legeai, F., Montagné, N., Poivet, E., Harry, M., Calatayud, P.-A., and Jacquin-Joly, E. (2013). Candidate Chemosensory Genes in the Stemborer *Sesamia nonagrioides*. *Int. J. Biol. Sci.* *9*, 481–495.
- Glaser, N., Gallot, A., Legeai, F., Harry, M., Kaiser, L., Le Ru, B., Calatayud, P.-A., and Jacquin-Joly, E. (2015). Differential expression of the chemosensory transcriptome in two populations of the stemborer *Sesamia nonagrioides*. *Insect Biochem. Mol. Biol.* *65*, 28–34.
- Gong, D.-P., Zhang, H.-J., Zhao, P., Xia, Q.-Y., and Xiang, Z.-H. (2009). The Odorant Binding Protein Gene Family from the Genome of Silkworm, *Bombyx mori*. *BMC Genomics* *10*, 332.
- González Audino, P., Vassena, C., Barrios, S., Zerba, E., and Picollo, M.I. (2004). Role of enhanced detoxication in a deltamethrin-resistant population of *Triatoma infestans* (Hemiptera, Reduviidae) from Argentina. *Mem. Inst. Oswaldo Cruz* *99*, 335–339.
- Gourbière, S., Dorn, P., Tripet, F., and Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* *108*, 190–202.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* *29*, 644–652.
- Gregorc, A., Evans, J.D., Scharf, M., and Ellis, J.D. (2012). Gene expression in honey bee (*Apis mellifera*) larvae exposed to pesticides and Varroa mites (*Varroa destructor*). *J. Insect Physiol.* *58*, 1042–1049.

- Grosjean, Y., Rytz, R., Farine, J.-P., Abuin, L., Cortot, J., Jefferis, G.S.X.E., and Benton, R. (2011). An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature* *478*, 236–240.
- Große-Wilde, E., Svatoš, A., and Krieger, J. (2006). A Pheromone-Binding Protein Mediates the Bombykol-Induced Activation of a Pheromone Receptor In Vitro. *Chem. Senses* *31*, 547–555.
- Gschloessl, B., Beyne, E., Audiot, P., Bourguet, D., and Streiff, R. (2013). De novo transcriptomic resources for two sibling species of moths: *Ostrinia nubilalis* and *O. scapularis*. *BMC Res. Notes* *6*, 73.
- Gu, S.-H., Wang, S.-P., Zhang, X.-Y., Wu, K.-M., Guo, Y.-Y., Zhou, J.-J., and Zhang, Y.-J. (2011). Identification and tissue distribution of odorant binding protein genes in the lucerne plant bug *Adelphocoris lineolatus* (Goeze). *Insect Biochem. Mol. Biol.* *41*, 254–263.
- Gu, S.-H., Wang, S.-Y., Zhang, X.-Y., Ji, P., Liu, J.-T., Wang, G.-R., Wu, K.-M., Guo, Y.-Y., Zhou, J.-J., and Zhang, Y.-J. (2012). Functional Characterizations of Chemosensory Proteins of the Alfalfa Plant Bug *Adelphocoris lineolatus* Indicate Their Involvement in Host Recognition. *PLoS ONE* *7*.
- Gu, S.-H., Wu, K.-M., Guo, Y.-Y., Pickett, J.A., Field, L.M., Zhou, J.-J., and Zhang, Y.-J. (2013). Identification of genes expressed in the sex pheromone gland of the black cutworm *Agrotis ipsilon* with putative roles in sex pheromone biosynthesis and transport. *BMC Genomics* *14*, 636.
- Guarneri, A.A., Lazzari, C., Xavier, A.A.P., Diotaiuti, L., and Lorenzo, M.G. (2003). The effect of temperature on the behaviour and development of *Triatoma brasiliensis*. *Physiol. Entomol.* *28*, 185–191.
- Guerenstein, P., and Guerin, P. (2001). Olfactory and behavioural responses of the blood-sucking bug *Triatoma infestans* to odours of vertebrate hosts. *J. Exp. Biol.* *204*, 585–597.
- Guerenstein, P.G., and Hildebrand, J.G. (2008). Roles and effects of environmental carbon dioxide in insect life. In *Annual Review of Entomology*, (Palo Alto: Annual Reviews), pp. 161–178.
- Guhl, F., Restrepo, M., Angulo, V.M., Antunes, C.M.F., Campbell-Lendrum, D., and Davies, C.R. (2005). Lessons from a national survey of Chagas disease transmission risk in Colombia. *Trends Parasitol.* *21*, 259–262.
- Guhl, F., Pinto, N., and Aguilera, G. (2009). Sylvatic triatominae: a new challenge in vector control transmission. *Mem. Inst. Oswaldo Cruz* *104*, 71–75.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* *59*, 307–321.
- Gunning, R.V., Devonshire, A.L., and Moores, G.D. (1995). Metabolism of Esfenvalerate by Pyrethroid-Susceptible and -Resistant Australian *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Pestic. Biochem. Physiol.* *51*, 205–213.
- Guo, S., and Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Mol. Biol. Evol.* *24*, 1198–1207.
- Guo, W., Wang, X., Ma, Z., Xue, L., Han, J., Yu, D., and Kang, L. (2011). CSP and Takeout Genes Modulate the Switch between Attraction and Repulsion during Behavioral Phase Change in the Migratory Locust. *PLoS Genet* *7*, e1001291.
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* *29*, 1072–1075.
- Gurevitz, J.M., Ceballos, L.A., Kitron, U., and Gurtler, R.E. (2006). Flight initiation of *Triatoma infestans* (Hemiptera : reduviidae) under natural climatic conditions. *J. Med. Entomol.* *43*, 143–150.
- Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* *28*, 503–510.
- Györgyi, T.K., Roby-Shemkovitz, A.J., and Lerner, M.R. (1988). Characterization and cDNA cloning of the pheromone-binding protein from the tobacco hornworm, *Manduca sexta*: a tissue-specific developmentally regulated protein. *Proc. Natl. Acad. Sci.* *85*, 9851–9855.
- Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series*, pp. 95–98.
- Hallam, E.A., Ho, M.G., and Carlson, J.R. (2004). The Molecular Basis of Odor Coding in the *Drosophila* Antenna. *Cell* *117*, 965–979.
- Harrop, T.W.R., Pearce, S.L., Daborn, P.J., and Batterham, P. (2014). Whole-Genome Expression Analysis in the Third Instar Larval Midgut of *Drosophila melanogaster*. *G3 GenesGenomesGenetics* *4*, 2197–2205.
- Harry, M., Galindez, I., and Cariou, M.L. (1992). Isozyme variability and differentiation between *Rhodnius prolixus*, *R. robustus* and *R. pictipes*, vectors of Chagas disease in Venezuela. *Med. Vet. Entomol.* *6*, 37–43.

- Harry, M. (1993a). Isozymic data question the specific status of some blood-sucking bugs of the genus *Rhodnius*, vectors of Chagas disease. *Trans. R. Soc. Trop. Med. Hyg.* *87*, 492–492.
- Harry, M. (1993b). Use of the median process of the pygophore in the identification of *Rhodnius nasutus*, *R. neglectus*, *R. prolixus* and *R. robustus* (Hemiptera: Reduviidae). *Ann. Trop. Med. Parasitol.* *87*, 277–282.
- Harry, M. (1994). Morphometric variability in the Chagas' disease vector *Rhodnius prolixus*. *Jpn. J. Genet.* *69*, 233–250.
- Harry, M., Galindez, I., and Cariou, M.L. (1992). Isozyme variability and differentiation between *Rhodnius prolixus*, *R. robustus* and *R. pictipes*, vectors of Chagas disease in Venezuela. *Med. Vet. Entomol.* *6*, 37–43.
- Hashimoto, K., and Schofield, C.J. (2012). Elimination of *Rhodnius prolixus* in Central America. *Parasit. Vectors* *5*, 45.
- Haznedaroglu, B.Z., Reeves, D., Rismani-Yazdi, H., and Peccia, J. (2012). Optimization of de novo transcriptome assembly from high-throughput short read sequencing data improves functional annotation for non-model organisms. *BMC Bioinformatics* *13*, 170.
- Hekmat-Scafe, D.S., Scafe, C.R., McKinney, A.J., and Tanouye, M.A. (2002). Genome-Wide Analysis of the Odorant-Binding Protein Gene Family in *Drosophila melanogaster*. *Genome Res.* *12*, 1357–1369.
- Henrichsen, C.N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H., and Reymond, A. (2009a). Segmental copy number variation shapes tissue transcriptomes. *Nat. Genet.* *41*, 424–429.
- Henrichsen, C.N., Chaignat, E., and Reymond, A. (2009b). Copy number variants, diseases and gene expression. *Hum. Mol. Genet.* *18*, R1–R8.
- Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M., and Zwiebel, L.J. (2002). G Protein-Coupled Receptors in *Anopheles gambiae*. *Science* *298*, 176–178.
- Hodges, T.K., Cosme, L.V., Athrey, G., Pathikonda, S., Takken, W., and Slotman, M.A. (2014). RETRACTED ARTICLE: Species-specific chemosensory gene expression in the olfactory organs of the malaria vector *Anopheles gambiae*. *BMC Genomics* *15*, 1089.
- Holloway, A.K., Lawniczak, M.K.N., Mezey, J.G., Begun, D.J., and Jones, C.D. (2007). Adaptive Gene Expression Divergence Inferred from Population Genomics. *PLoS Genet* *3*, e187.
- Hooper, A.M., Dufour, S., He, X., Muck, A., Zhou, J.-J., Almeida, R., Field, L.M., Svatoš, A., and Pickett, J.A. (2009). High-throughput ESI-MS analysis of binding between the *Bombyx mori* pheromone-binding protein BmorPBP1, its pheromone components and some analogues. *Chem. Commun.* 5725–5727.
- Hornett, E.A., and Wheat, C.W. (2012). Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *Bmc Genomics* *13*, 361.
- Hua, J.-F., Zhang, S., Cui, J.-J., Wang, D.-J., Wang, C.-Y., Luo, J.-Y., and Lv, L.-M. (2012). Identification and Binding Characterization of Three Odorant Binding Proteins and One Chemosensory Protein from *Apolygus lucorum* (Meyer-Dur). *J. Chem. Ecol.* *38*, 1163–1170.
- Hua, J.-F., Zhang, S., Cui, J.-J., Wang, D.-J., Wang, C.-Y., Luo, J.-Y., Lv, L.-M., and Ma, Y. (2013). Functional characterizations of one odorant binding protein and three chemosensory proteins from *Apolygus lucorum* (Meyer-Dur) (Hemiptera: Miridae) legs. *J. Insect Physiol.* *59*, 690–696.
- Huang, X., and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Res.* *9*, 868–877.
- Hull, J.J., Geib, S.M., Fabrick, J.A., and Brent, C.S. (2013). Sequencing and De Novo Assembly of the Western Tarnished Plant Bug (*Lygus hesperus*) Transcriptome. *PLoS ONE* *8*, e55105.
- Hull, J.J., Perera, O.P., and Snodgrass, G.L. (2014). Cloning and expression profiling of odorant-binding proteins in the tarnished plant bug, *Lygus lineolaris*. *Insect Mol. Biol.* *23*, 78–97.
- Hypša, V., Tietz, D.F., Zrzavý, J., Rego, R.O.M., Galvao, C., and Jurberg, J. (2002). Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade. *Mol. Phylogenet. Evol.* *23*, 447–457.
- Ibarra-Laclette, E., Méndez-Bravo, A., Pérez-Torres, C.A., Albert, V.A., Mockaitis, K., Kilaru, A., López-Gómez, R., Cervantes-Luevano, J.I., and Herrera-Estrella, L. (2015). Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genomics* *16*, 599.
- Ishida, Y., Ishibashi, J., and Leal, W.S. (2013). Fatty Acid Solubilizer from the Oral Disk of the Blowfly. *PLoS ONE* *8*.
- Jacquin-Joly, E., and Merlin, C. (2004). INSECT OLFACTORY RECEPTORS: CONTRIBUTIONS OF MOLECULAR BIOLOGY TO CHEMICAL ECOLOGY. *J. Chem. Ecol.* *30*, 2359–2397.

- Jacquín-Joly, E., Vogt, R.G., François, M.-C., and Meillour, P.N.-L. (2001). Functional and Expression Pattern Analysis of Chemosensory Proteins Expressed in Antennae and Pheromonal Gland of *Mamestra brassicae*. *Chem. Senses* 26, 833–844.
- Ji, P., Gu, S.-H., Liu, J.-T., Zhu, X.-Q., Guo, Y.-Y., Zhou, J.-J., and Zhang, Y.-J. (2013). Identification and expression profile analysis of odorant-binding protein genes in *Apolygus lucorum* (Hemiptera: Miridae). *Appl. Entomol. Zool.* 48, 301–311.
- Kahn, A. (1985). Gènes, pseudogènes, familles multigéniques et évolution. *Médecine/sciences* 1, 214.
- Kaissling, K.-E. (2009). Olfactory perireceptor and receptor events in moths: a kinetic model revised. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 195, 895–922.
- Karatolos, N., Pauchet, Y., Wilkinson, P., Chauhan, R., Denholm, I., Gorman, K., Nelson, D.R., Bass, C., French-Constant, R.H., and Williamson, M.S. (2011). Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. *BMC Genomics* 12, 56.
- Karlson, P., and Luscher, M. (1959). Pheromones: a new term for a class of biologically active substances. *Nature* 183, 55–56.
- Kato, H., Jochim, R.C., Gomez, E.A., Sakoda, R., Iwata, H., Valenzuela, J.G., and Hashiguchi, Y. (2010). A repertoire of the dominant transcripts from the salivary glands of the blood-sucking bug, *Triatoma dimidiata*, a vector of Chagas disease. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 10, 184–191.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D., Hudson, P., Jolles, A., Jones, K.E., Mitchell, C.E., et al. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468, 647–652.
- Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci.* 107, 12168–12173.
- Knudsen, B., Knudsen, T., Flensburg, M., Sandmann, H., Heltzen, M., Andersen, A., Dickenson, M., Bardram, K., Steffensen, P., and Monsted, S. (2007). CLC Genomics Workbench. Version 5.5.
- Kondrashov, F.A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. Lond. B Biol. Sci.* 279, 5048–5057.
- Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700.
- Krieger, J., Raming, K., and Breer, H. (1991). Cloning of genomic and complementary DNA encoding insect pheromone binding proteins: evidence for microdiversity. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* 1088, 277–284.
- Krieger, J., von Nickisch-Rosenegk, E., Mameli, M., Pelosi, P., and Breer, H. (1996). Binding proteins from the antennae of *Bombyx mori*. *Insect Biochem. Mol. Biol.* 26, 297–307.
- Krieger, J., Klink, O., Mohl, C., Raming, K., and Breer, H. (2003). A candidate olfactory receptor subtype highly conserved across different insect orders. *J. Comp. Physiol. A* 189, 519–526.
- Kulmuni, J., Wurm, Y., and Pamilo, P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity* 110, 538–547.
- Larsson, M.C., Domingos, A.I., Jones, W.D., Chiappe, M.E., Amrein, H., and Vossell, L.B. (2004). Or83b Encodes a Broadly Expressed Odorant Receptor Essential for *Drosophila* Olfaction. *Neuron* 43, 703–714.
- Latorre Estivalis, J.M., Omondi, B.A., DeSouza, O., Oliveira, I.H.R., Ignell, R., and Lorenzo, M.G. (2015a). Molecular basis of peripheral olfactory plasticity in *Rhodnius prolixus*, a Chagas disease vector. *Chem. Ecol.* 3, 74.
- Latorre-Estivalis, J.M., de Oliveira, E.S., Beiral Esteves, B., Santos Guimarães, L., Neves Ramos, M., and Lorenzo, M.G. (2015b). Patterns of expression of odorant receptor genes in a Chagas disease vector. *Insect Biochem. Mol. Biol.*
- Laughlin, J.D., Ha, T.S., Jones, D.N.M., and Smith, D.P. (2008). Activation of Pheromone-Sensitive Neurons Is Mediated by Conformational Activation of Pheromone-Binding Protein. *Cell* 133, 1255–1265.
- Lazzari, C.R. (1992). Circadian organization of locomotion activity in the haematophagous bug *Triatoma infestans*. *J. Insect Physiol.* 38, 895–903.

- Lazzari, C.R., Pereira, M.H., Lorenzo, M.G., Lazzari, C.R., Pereira, M.H., and Lorenzo, M.G. (2013). Behavioural biology of Chagas disease vectors. *Mem. Inst. Oswaldo Cruz* 108, 34–47.
- Leal, W.S., Nikonova, L., and Peng, G. (1999). Disulfide structure of the pheromone binding protein from the silkworm moth, *Bombyx mori*. *FEBS Lett.* 464, 85–90.
- Leal, W.S., Parra-Pedrazzoli, A.L., Kaissling, K.-E., Morgan, T.I., Zalom, F.G., Pesak, D.J., Dundulis, E.A., Burks, C.S., and Higbee, B.S. (2005). Unusual pheromone chemistry in the navel orangeworm: novel sex attractants and a behavioral antagonist. *Naturwissenschaften* 92, 139–146.
- Leal, W.S., Choo, Y.-M., Xu, P., da Silva, C.S.B., and Ueira-Vieira, C. (2013). Differential expression of olfactory genes in the southern house mosquito and insights into unique odorant receptor gene isoforms. *Proc. Natl. Acad. Sci. U. S. A.*
- Leary, G.P., Allen, J.E., Bunger, P.L., Luginbill, J.B., Linn, C.E., Macallister, I.E., Kavanaugh, M.P., and Wanner, K.W. (2012). Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14081–14086.
- Lee, B.Y., Bacon, K.M., Bottazzi, M.E., and Hotez, P.J. (2013). Global economic burden of Chagas disease: a computational simulation model. *Lancet Infect. Dis.* 13, 342–348.
- Lee, Y., Moon, S.J., and Montell, C. (2009). Multiple gustatory receptors required for the caffeine response in *Drosophila*. *Proc. Natl. Acad. Sci.* 106, 4495–4500.
- Lee, Y., Kang, M.J., Shim, J., Cheong, C.U., Moon, S.J., and Montell, C. (2012). Gustatory Receptors Required for Avoiding the Insecticide l-Canavanine. *J. Neurosci.* 32, 1429–1435.
- Lehane, M.J., and Schofield, C.J. (1982). Flight initiation in *Triatoma infestans* (Klug) (Hemiptera: Reduviidae). *Bull. Entomol. Res.* 72, 497–510.
- Lent, H., and Wygodzinsky, P. (1979). Revision of the triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Revisión de los triatominae (Hemiptera, Reduviidae) y su significado como vectores del mal de Chagas.* *Bull Am Mus Nat Hist* 163, 123–520.
- Lescure, F.-X., Canestri, A., Melliez, H., Jauréguiberry, S., Develoux, M., Dorent, R., Guiard-Schmid, J.-B., Bonnard, P., Ajana, F., Rolla, V., et al. (2008). Chagas Disease, France. *Emerg. Infect. Dis.* 14, 644–649.
- Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128.
- Levy, M.Z., Tustin, A., Castillo-Neyra, R., Mabud, T.S., Levy, K., Barbu, C.M., Quispe-Machaca, V.R., Ancca-Juarez, J., Borrini-Mayori, K., Naquira-Velarde, C., et al. (2015). Bottlenecks in domestic animal populations can facilitate the emergence of *Trypanosoma cruzi*, the aetiological agent of Chagas disease. *Proc R Soc B* 282, 20142807.
- Lewis, M.D., Llewellyn, M.S., Yeo, M., Acosta, N., Gaunt, M.W., and Miles, M.A. (2011). Recent, Independent and Anthropogenic Origins of *Trypanosoma cruzi* Hybrids. *PLoS Negl Trop Dis* 5, e1363.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, H., Zhang, H., Guan, R., and Miao, X. (2013). Identification of differential expression genes associated with host selection and adaptation between two sibling insect species by transcriptional profile analysis. *BMC Genomics* 14, 582.
- Li, S., Picimbon, J.-F., Ji, S., Kan, Y., Chuanling, Q., Zhou, J.-J., and Pelosi, P. (2008). Multiple functions of an odorant-binding protein in the mosquito *Aedes aegypti*. *Biochem. Biophys. Res. Commun.* 372, 464–468.
- Lin, Y., Jin, T., Zeng, L., and Lu, Y. (2012). Cuticular penetration of β -cypermethrin in insecticide-susceptible and resistant strains of *Bactrocera dorsalis*. *Pestic. Biochem. Physiol.* 103, 189–193.
- Liu, G.X., Xuan, N., Chu, D., Xie, H.Y., Fan, Z.X., Bi, Y.P., Picimbon, J.-F., Qin, Y.C., Zhong, S.T., Fa Li, Y., et al. (2014). BIOTYPE EXPRESSION AND INSECTICIDE RESPONSE OF *Bemisia tabaci* CHEMOSENSORY PROTEIN-1. *Arch. Insect Biochem. Physiol.* 85, 137–151.
- Liu, R., He, X., Lehane, S., Lehane, M., Hertz-Fowler, C., Berriman, M., Field, L.M., and Zhou, J.-J. (2012a). Expression of chemosensory proteins in the tsetse fly *Glossina morsitans morsitans* is related to female host-seeking behaviour. *Insect Mol. Biol.* 21, 41–48.
- Liu, S., Chougule, N.P., Vijayendran, D., and Bonning, B.C. (2012b). Deep Sequencing of the Transcriptomes of Soybean Aphid and Associated Endosymbionts. *PLoS ONE* 7, e45161.

- Lopez, G., Moreno, J., Lopez, G., and Moreno, J. (1995). Genetic variability and differentiation between populations of *Rhodnius prolixus* and *R. pallescens*, vectors of Chagas' Disease in Colombia. *Mem. Inst. Oswaldo Cruz* *90*, 353–357.
- Lopez-Antunano, F.J., Rangel-Flores, H., and Ramos, C. (2000). Diagnosis of Chagas' disease. *Rev. Latinoam. Microbiol.-Mex.-* *42*, 121–130.
- López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* *9*, 583–593.
- Lorenzo, M.G., and Lazzari, C.R. (1996). The spatial pattern of defaecation in *Triatoma infestans* and the role of faeces as a chemical mark of the refuge. *J. Insect Physiol.* *42*, 903–907.
- Lorenzo, M.G., and Lazzari, C.R. (1998). Activity pattern in relation to refuge exploitation and feeding in *Triatoma infestans* (Hemiptera: Reduviidae). *Acta Trop.* *70*, 163–170.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*.
- Luz, C., Fargues, J., and Grunewald, J. (1999). Development of *Rhodnius prolixus* (Hemiptera: Reduviidae) under Constant and Cyclic Conditions of Temperature and Humidity. *Mem. Inst. Oswaldo Cruz* *94*, 403–409.
- Lyman, D.F., Monteiro, F.A., Escalante, A.A., Cordon-Rosales, C., Wesson, D.M., Dujardin, J.P., and Beard, C.B. (1999). Mitochondrial DNA sequence variation among triatomine vectors of Chagas' disease. *Am. J. Trop. Med. Hyg.* *60*, 377–386.
- Mace, G., Masundire, H., and Baillie, J. (2005). Biodiversity. In *Ecosystems and Human Well-Being: Current State and Trends*, (Island Press),.
- Madden, T. (2003). The BLAST Sequence Analysis Tool.
- Maia Da Silva, F., Junqueira, A.C.V., Campaner, M., Rodrigues, A.C., Crisante, G., Ramirez, L.E., Caballero, Z.C.E., Monteiro, F.A., Coura, J.R., Añez, N., et al. (2007). Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. *Mol. Ecol.* *16*, 3361–3373.
- Maleszka, J., Forêt, S., Saint, R., and Maleszka, R. (2007). RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*). *Dev. Genes Evol.* *217*, 189–196.
- Malpel, S., Merlin, C., François, M.-C., and Jacquin-Joly, E. (2008). Molecular identification and characterization of two new Lepidoptera chemoreceptors belonging to the *Drosophila melanogaster* OR83b family. *Insect Mol. Biol.* *17*, 587–596.
- Mameli, M., Tuccini, A., Mazza, M., Petacchi, R., and Pelosi, P. (1996). Soluble proteins in chemosensory organs of phasids. *Insect Biochem. Mol. Biol.* *26*, 875–882.
- Mamidala, P., Wijeratne, A.J., Wijeratne, S., Kornacker, K., Sudhamalla, B., Rivera-Vega, L.J., Hoelmer, A., Meulia, T., Jones, S.C., and Mittapalli, O. (2012). RNA-Seq and molecular docking reveal multi-level pesticide resistance in the bed bug. *BMC Genomics* *13*, 6.
- Manrique, G., Vitta, A.C.R., Ferreira, R.A., Zani, C.L., Unelius, C.R., Lazzari, C.R., Diotaiuti, L., and Lorenzo, M.G. (2006). Chemical Communication in Chagas Disease Vectors. Source, Identity, and Potential Function of Volatiles Released by the Metasternal and Brindley's Glands of *Triatoma infestans* Adults. *J. Chem. Ecol.* *32*, 2035–2052.
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* *27*, 764–770.
- Marchant, A., Mougél, F., Almeida, C., Jacquin-Joly, E., Costa, J., and Harry, M. (2015a). De novo transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease. *Genetica* *143*, 225–239.
- Marchant, A., Mougél, F., Mendonça, V., Quartier, M., Jacquin-Joly, E., da Rosa, A., Petit, E., and Harry, M. (2015b). Comparing de novo and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.*
- Marcili, A., Lima, L., Cavazzana, M., Junqueira, A.C.V., Veludo, H.H., Maia Da Silva, F., Campaner, M., Paiva, F., Nunes, V.L.B., and Teixeira, M.M.G. (2009). A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology* *136*, 641–655.

- Marcombe, S., Mathieu, R.B., Pocquet, N., Riaz, M.-A., Poupardin, R., Sélis, S., Darriet, F., Reynaud, S., Yébakima, A., Corbel, V., et al. (2012). Insecticide Resistance in the Dengue Vector *Aedes aegypti* from Martinique: Distribution, Mechanisms and Relations with Environmental Factors. *PLoS ONE* 7, e30989.
- Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24, 133–141.
- Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. *Cell. Mol. Life Sci.* 67, 569–579.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- Marinotti, O., Ngo, T., Kojin, B.B., Chou, S.-P., Nguyen, B., Juhn, J., Carballar-Lejarazú, R., Marinotti, P.N., Jiang, X., Walter, M.F., et al. (2014). Integrated proteomic and transcriptomic analysis of the *Aedes aegypti* eggshell. *BMC Dev. Biol.* 14, 15.
- Márquez, E.J., and Saldamando-Benjumea, C.I. (2013). *Rhodnius prolixus* and *Rhodnius robustus*-like (Hemiptera, Reduviidae) wing asymmetry under controlled conditions of population density and feeding frequency. *J. Biosci.* 38, 549–560.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, pp. 10–12.
- Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12, 671–682.
- Martin, J., Bruno, V.M., Fang, Z., Meng, X., Blow, M., Zhang, T., Sherlock, G., Snyder, M., and Wang, Z. (2010). Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* 11, 663.
- Martínez-Barnette, J., Gómez-Barreto, R.E., Ovillla-Muñoz, M., Téllez-Sosa, J., López, D.E.G., Dinglasan, R.R., Mohien, C.U., MacCallum, R.M., Redmond, S.N., Gibbons, J.G., et al. (2012). Transcriptome of the adult female malaria mosquito vector *Anopheles albimanus*. *BMC Genomics* 13, 207.
- May-Concha, I., Rojas, J.C., Cruz-López, L., Millar, J.G., and Ramsey, J.M. (2013). Volatile compounds emitted by *Triatoma dimidiata*, a vector of Chagas disease: chemical analysis and behavioural evaluation. *Med. Vet. Entomol.* 27, 165–174.
- McBride, C.S., Baier, F., Omondi, A.B., Spitzer, S.A., Lutomiah, J., Sang, R., Ignell, R., and Vosshall, L.B. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* 515, 222–U151.
- McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405.
- McKenna, M.P., Hekmat-Scafe, D.S., Gaines, P., and Carlson, J.R. (1994). Putative *Drosophila* pheromone-binding proteins expressed in a subregion of the olfactory system. *J. Biol. Chem.* 269, 16340–16347.
- Medeiros, M.N., Logullo, R., Ramos, I.B., Sorgine, M.H.F., Paiva-Silva, G.O., Mesquita, R.D., Machado, E.A., Coutinho, M.A., Masuda, H., Capurro, M.L., et al. (2011). Transcriptome and gene expression profile of ovarian follicle tissue of the triatomine bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 41, 823–831.
- Medone, P., Ceccarelli, S., Parham, P.E., Figuera, A., and Rabinovich, J.E. (2015). The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130560.
- Mendonça, V.J., Silva, M.T.A. da, Araújo, R.F. de, Júnior, J.M., Júnior, M.B., Almeida, C.E., Costa, J., Graminha, M.A.S., Cicarelli, R.M.B., and Rosa, J.A. da (2009). Phylogeny of *Triatoma sherlocki* (Hemiptera: Reduviidae: Triatominae) Inferred from Two Mitochondrial Genes Suggests Its Location Within the *Triatoma brasiliensis* Complex. *Am. J. Trop. Med. Hyg.* 81, 858–864.
- Meneguetti, D.U. de O., Tojal, S.D., Miranda, P.R.M. de, Rosa, J.A. da, Camargo, L.M.A., Meneguetti, D.U. de O., Tojal, S.D., Miranda, P.R.M. de, Rosa, J.A. da, and Camargo, L.M.A. (2015). First report of *Rhodnius montenegrensis* (Hemiptera, Reduviidae, Triatominae) in the State of Acre, Brazil. *Rev. Soc. Bras. Med. Trop.* 48, 471–473.
- Mercier, J.-L., Errard, C., and Frérot, B. (2013). L'identité chimique des insectes. In *Des Insectes et Des Plantes*.
- Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F.A., Minx, P., Spieth, J., Carvalho, A.B., Panzera, F., Lawson, D., et al. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci.* 201506226.

- Messenger, L.A., Miles, M.A., and Bern, C. (2015). Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. *Expert Rev. Anti Infect. Ther.* *13*, 995–1029.
- Metzker, M.L. (2010). Sequencing technologies — the next generation. *Nat. Rev. Genet.* *11*, 31–46.
- Meunier, N., Belgacem, Y.H., and Martin, J.-R. (2007). Regulation of feeding behaviour and locomotor activity by takeout in *Drosophila*. *J. Exp. Biol.* *210*, 1424–1434.
- Miles, M.A., Arias, J.R., and Souza, A. (1983). Chagas’ disease in the Amazon Basin: V. Periurban palms as habitats of *Rhodnius robustus* and *Rhodnius pictipes*-Triatomine vectors of Chagas’ disease. *Mem. Inst. Oswaldo Cruz* *78*, 391–398.
- Miles, M.A., Feliciangeli, M.D., and Arias, A.R. de (2003). American trypanosomiasis (Chagas’ disease) and the role of molecular epidemiology in guiding control strategies. *BMJ* *326*, 1444–1448.
- Minoli, S.A., and Lazzari, C.R. (2006). Take-off activity and orientation of triatomines (Heteroptera: Reduviidae) in relation to the presence of artificial lights. *Acta Trop.* *97*, 324–330.
- Minoli, S., Palottini, F., and Manrique, G. (2013). The main component of an alarm pheromone of kissing bugs plays multiple roles in the cognitive modulation of the escape response. *Front. Behav. Neurosci.* *7*.
- Miyamoto, T., and Amrein, H. (2008). Suppression of male courtship by a *Drosophila* pheromone receptor. *Nat. Neurosci.* *11*, 874–876.
- Monteiro, F.A., Pérez, R., Panzera, F., Dujardin, J.-P., Galvão, C., Rocha, D., Noireau, F., Schofield, C., and Beard, C.B. (1999). Mitochondrial DNA variation of *Triatoma infestans* populations and its implication on the specific status of *T. melanosoma*. *Mem. Inst. Oswaldo Cruz* *94*, 229–238.
- Monteiro, F.A., Wesson, D.M., Dotson, E.M., Schofield, C.J., and Beard, C.B. (2000). Phylogeny and molecular taxonomy of the Rhodniini derived from mitochondrial and nuclear DNA sequences. *Am. J. Trop. Med. Hyg.* *62*, 460–465.
- Monteiro, F.A., Barrett, T.V., Fitzpatrick, S., Cordon-Rosales, C., Feliciangeli, D., and Beard, C.B. (2003). Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Mol. Ecol.* *12*, 997–1006.
- Monteiro, F.A., Donnelly, M.J., Beard, C.B., and Costa, J. (2004). Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in Northeast Brazil. *Mol. Phylogenet. Evol.* *32*, 46–56.
- Monteiro, W.M., Magalhães, L.K., Santana Filho, F.S., Borborema, M., Silveira, H., and Barbosa, M. das G.V. (2010). *Trypanosoma cruzi* TcIII / Z3 genotype as agent of an outbreak of Chagas disease in the Brazilian Western Amazonia. *Trop. Med. Int. Health* *15*, 1049–1051.
- Moon, S.J., Köttgen, M., Jiao, Y., Xu, H., and Montell, C. (2006). A Taste Receptor Required for the Caffeine Response In Vivo. *Curr. Biol.* *16*, 1812–1817.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* *5*, 621–628.
- Mosbah, A., Campanacci, V., Lartigue, A., Tegoni, M., Cambillau, C., and Darbon, H. (2003). Solution structure of a chemosensory protein from the moth *Mamestra brassicae*. *Biochem. J.* *369*, 39–44.
- Mougabure-Cueto, G., and Picollo, M.I. (2015). Insecticide resistance in vector Chagas disease: Evolution, mechanisms and management. *Acta Trop.* *149*, 70–85.
- Mundry, M., Bornberg-Bauer, E., Sammeth, M., and Feulner, P.G.D. (2012). Evaluating Characteristics of De Novo Assembly Software on 454 Transcriptome Data: A Simulation Approach. *PLoS ONE* *7*, e31410.
- Nei, M., and Rooney, A.P. (2005). Concerted and Birth-and-Death Evolution of Multigene Families. *Annu. Rev. Genet.* *39*, 121–152.
- Nkya, T.E., Akhouayri, I., Kisinza, W., and David, J.-P. (2013). Impact of environment on mosquito response to pyrethroid insecticides: Facts, evidences and prospects. *Insect Biochem. Mol. Biol.* *43*, 407–416.
- Noireau, F., and Dujardin, J.-P. (2001). Flight and nutritional status of sylvatic *Triatoma sordida* and *Triatoma guayanae*. *Mem. Inst. Oswaldo Cruz* *96*, 385–389.
- Noireau, F., Flores, R., Gutierrez, T., Abad-Franch, F., Flores, E., and Vargas, F. (2000). Natural ecotopes of *Triatoma infestans* dark morph and other sylvatic triatomines in the Bolivian Chaco. *Trans. R. Soc. Trop. Med. Hyg.* *94*, 23–27.
- Noireau, F., Diosque, P., and Jansen, A.M. (2009). *Trypanosoma cruzi*: adaptation to its vectors and its hosts. *Vet. Res.* *40*.
- Nomura, A., Kawasaki, K., Kubo, T., and Natori, S. (1992). Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *Int. J. Dev. Biol.* *36*, 391–398.

- Nomura Kitabayashi, A., Arai, T., Kubo, T., and Natori, S. (1998). Molecular cloning of cDNA for p10, a novel protein that increases in the regenerating legs of *Periplaneta americana* (American cockroach). *Insect Biochem. Mol. Biol.* *28*, 785–790.
- O’Neil, S.T., and Emrich, S.J. (2013). Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* *14*, 465.
- Ohta, T. (2013). *Evolution and Variation of Multigene Families* (Springer Science & Business Media).
- Okasha, A.Y.K. (1964). Effects of High Temperature in *Rhodnius prolixus* (Stål). *Nature* *204*, 1221–1222.
- OMS (2011). Agir pour réduire l’impact mondial des maladies tropicales négligées : premier rapport de l’OMS sur les maladies tropicales négligées.
- OMS | Maladie de Chagas (trypanosomiase américaine).
- Orihuela, P.L.S., Vassena, C.V., Zerba, E.N., and Picollo, M.I. (2008). Relative Contribution of Monooxygenase and Esterase to Pyrethroid Resistance in *Triatoma infestans* (Hemiptera: Reduviidae) from Argentina and Bolivia. *J. Med. Entomol.* *45*, 298–306.
- Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* *11*, 220.
- Parham, P.E., Waldock, J., Christophides, G.K., Hemming, D., Agosto, F., Evans, K.J., Fefferman, N., Gaff, H., Gumel, A., LaDeau, S., et al. (2015). Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *370*, 20130551.
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* *23*, 1061–1067.
- Paszkiewicz, K., and Studholme, D.J. (2010). De novo assembly of short sequence reads. *Brief. Bioinform.* *11*, 457–472.
- Patterson, B.D. (1994). Accumulating Knowledge on the Dimensions of Biodiversity: Systematic Perspectives on Neotropical Mammals. *Biodivers. Lett.* *2*, 79–86.
- De Paula, A.S., Diotaiuti, L., and Galvão, C. (2007). Systematics and biogeography of *Rhodniini* (Heteroptera: Reduviidae: Triatominae) based on 16S mitochondrial rDNA sequences. *J. Biogeogr.* *34*, 699–712.
- Pavan, M.G., Mesquita, R.D., Lawrence, G.G., Lazoski, C., Dotson, E.M., Abubucker, S., Mitreva, M., Randall-Maher, J., and Monteiro, F.A. (2013). A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Infect. Genet. Evol.* *14*, 426–433.
- Pelletier, J., Guidolin, A., Syed, Z., Cornel, A.J., and Leal, W.S. (2010). Knockdown of a Mosquito Odorant-binding Protein Involved in the Sensitive Detection of Oviposition Attractants. *J. Chem. Ecol.* *36*, 245–248.
- Pelosi, P., Calvello, M., and Ban, L. (2005). Diversity of Odorant-binding Proteins and Chemosensory Proteins in Insects. *Chem. Senses* *30*, i291–i292.
- Pelosi, P., Zhou, J.-J., Ban, L.P., and Calvello, M. (2006). Soluble proteins in insect chemical communication. *Cell. Mol. Life Sci. CMLS* *63*, 1658–1676.
- Pelosi, P., Iovinella, I., Felicioli, A., and Dani, F.R. (2014). Soluble proteins of chemical communication: an overview across arthropods. *Front. Physiol.* *5*.
- Pentreath, V.W. (1995). Trypanosomiasis and the nervous system: Pathology and immunology. *Trans. R. Soc. Trop. Med. Hyg.* *89*, 9–15.
- Pereira, J., Dujardin, J.P., Salvatella, R., and Tibayrenc, M. (1996). Abstract of article: Enzymatic variability and phylogenetic relatedness among *Triatoma infestans*, *T. platensis*, *T. delpontei* and *T. rubrovaria*. *Heredity* *77*, 47–54.
- Pereira, K.S., Sciimidt, F.L., Guaraldo, A.M.A., Franco, R.M.B., Dias, V.L., and Passos, L.A.C. (2009). Chagas’ Disease as a Foodborne Illness. *J. Food Prot.* *72*, 441–446.
- Pessoa, G.C.D., Trevizani, N.A.B., Dias, L.S., Bezerra, C.M. de, Melo, B.V. de, Diotaiuti, L., Pessoa, G.C.D., Trevizani, N.A.B., Dias, L.S., Bezerra, C.M. de, et al. (2015). Toxicological profile of deltamethrin in *Triatoma brasiliensis* (Hemiptera: Reduviidae) in State of Ceará, Northeastern Brazil. *Rev. Soc. Bras. Med. Trop.* *48*, 39–43.
- Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* *8*, 785–786.
- Pfennig, D.W., Wund, M.A., Snell-Rood, E.C., Cruickshank, T., Schlichting, C.D., and Moczek, A.P. (2010). Phenotypic plasticity’s impacts on diversification and speciation. *Trends Ecol. Evol.* *25*, 459–467.
- Picimbon, J.-F., Dietrich, K., Krieger, J., and Breer, H. (2001). Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). *Insect Biochem. Mol. Biol.* *31*, 1173–1181.

- Pikielny, C.W., Hasan, G., Rouyer, F., and Rosbash, M. (1994). Members of a family of drosophila putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* 12, 35–49.
- Pillonel, J., Brouard, C., Laperche, S., Barin, F., Bernillon, P., and de Valk, H. (2009). Estimation quantitative du risque de contamination d'un don de sang par des agents infectieux. *Transfus. Clin. Biol.* 16, 138–145.
- Pinto, C.M., Kalko, E.K.V., Cottontail, I., Wellingshausen, N., and Cottontail, V.M. (2012). TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. *Infect. Genet. Evol.* 12, 1328–1332.
- Pitts, R.J., Rinker, D.C., Jones, P.L., Rokas, A., and Zwiebel, L.J. (2011). Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics* 12, 271.
- Poivet, E., Gallot, A., Montagné, N., Glaser, N., Legeai, F., and Jacquín-Joly, E. (2013). A Comparison of the Olfactory Gene Repertoires of Adults and Larvae in the Noctuid Moth *Spodoptera littoralis*. *PLoS ONE* 8, e60263.
- Pontes, G.B., Bohman, B., Unelius, C.R., and Lorenzo, M.G. (2008). Metasternal Gland Volatiles and Sexual Communication in the Triatomine Bug, *Rhodnius prolixus*. *J. Chem. Ecol.* 34, 450–457.
- Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Prata, A. (2001). Clinical and epidemiological aspects of Chagas disease. *Lancet Infect. Dis.* 1, 92–100.
- Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., Field, L.M., Millar, N.S., Williamson, M.S., and Bass, C. (2010). Amplification of a Cytochrome P450 Gene Is Associated with Resistance to Neonicotinoid Insecticides in the Aphid *Myzus persicae*. *PLoS Genet* 6, e1000999.
- Quartier, M. (2011). Etude de l'impact de l'anthropisation sur l'écologie évolutive des vecteurs de la maladie de Chagas : cas de trois communautés du Tapajos, Amazonie brésilienne. Université Paris-Est, Université de Neuchâtel.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rajpurohit, S., Oliveira, C.C., Etges, W.J., and Gibbs, A.G. (2013). Functional genomic and phenotypic responses to desiccation in natural populations of a desert drosophilid. *Mol. Ecol.* 22, 2698–2715.
- Raming, K., Krieger, J., and Breer, H. (1989). Molecular cloning of an insect pheromone-binding protein. *FEBS Lett.* 256, 215–218.
- Ramírez, J.D., Tapia-Calle, G., Muñoz-Cruz, G., Poveda, C., Rendón, L.M., Hincapié, E., and Guhl, F. (2014). Trypanosome species in neo-tropical bats: Biological, evolutionary and epidemiological implications. *Infect. Genet. Evol.* 22, 250–256.
- Rassi, A., Rassi, A., and Marin-Neto, J.A. (2010). Chagas disease. *The Lancet* 375, 1388–1402.
- Rau, A., Maugis-Rabusseau, C., Martin-Magniette, M.-L., and Celeux, G. (2015). Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. *Bioinformatics* 31, 1420–1427.
- Ren, X., Liu, T., Dong, J., Sun, L., Yang, J., Zhu, Y., and Jin, Q. (2012). Evaluating de Bruijn Graph Assemblers on 454 Transcriptomic Data. *PLoS ONE* 7, e51188.
- Rendón, L.M., Guhl, F., Cordovez, J.M., Erazo, D., Rendón, L.M., Guhl, F., Cordovez, J.M., and Erazo, D. (2015). New scenarios of *Trypanosoma cruzi* transmission in the Orinoco region of Colombia. *Mem. Inst. Oswaldo Cruz* 110, 283–288.
- Ribeiro, J.M., Andersen, J., Silva-Neto, M.A., Pham, V., Garfield, M., and Valenzuela, J. (2004). Exploring the salome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 34, 61–79.
- Ribeiro, J.M.C., Genta, F.A., Sorgine, M.H.F., Logullo, R., Mesquita, R.D., Paiva-Silva, G.O., Majerowicz, D., Medeiros, M., Koerich, L., Terra, W.R., et al. (2014). An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, *Rhodnius prolixus*. *PLoS Negl Trop Dis* 8, e2594.
- Richer, W., Kengne, P., Cortez, M.R., Perrineau, M.M., Cohuet, A., Fontenille, D., and Noireau, F. (2007). Active dispersal by wild *Triatoma infestans* in the Bolivian Andes. *Trop. Med. Int. Health* 12, 759–764.
- Rinker, D.C., Pitts, R.J., Zhou, X., Suh, E., Rokas, A., and Zwiebel, L.J. (2013). Blood meal-induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. *Proc. Natl. Acad. Sci.* 110, 8260–8265.
- Robertson, H.M., and Wanner, K.W. (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Res.* 16, 1395–1403.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., et al. (2010). De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912.

- Robertson, H.M., Warr, C.G., and Carlson, J.R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* *100*, 14537–14542.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.
- Romaña, C.A., Pizarro, J.C.N., Rodas, E., and Guilbert, E. (1999). Palm trees as ecological indicators of risk areas for Chagas disease. *Trans. R. Soc. Trop. Med. Hyg.* *93*, 594–595.
- Romaña, C.A., Brunstein, D., Collin-Delavaud, A., Sousa, O., and Ortega-Barria, E. (2003). Public policies of development in Latin America and Chagas' disease. *The Lancet* *362*, 579.
- Rosa, J.A., Rocha, C.S., Gardim, S., Pinto, M.C., Mendonca, V.J., Ferreira Filho, J.C.R., Carvalho, E.O.C., Camargo, L.M.A., Oliveira, J., and Nascimento, J.D. (2012). Description of *Rhodnius montenegrensis* n. sp. (Hemiptera: Reduviidae: Triatominae) from the state of Rondônia, Brazil. *Zootaxa* *3478*, 62–76.
- Roush, R., and Tabashnik, B.E. (2012). *Pesticide Resistance in Arthropods* (Springer Science & Business Media).
- Roux, E., De Fátima Venâncio, A., Girres, J.-F., and Romaña, C.A. (2011). Spatial patterns and eco-epidemiological systems – part II: characterising spatial patterns of the occurrence of the insect vectors of Chagas disease based on remote sensing and field data. *Geospatial Health* *6*, 53.
- Sakurai, T., Nakagawa, T., Mitsuno, H., Mori, H., Endo, Y., Tanoue, S., Yasukochi, Y., Touhara, K., and Nishioka, T. (2004). Identification and functional characterization of a sex pheromone receptor in the silkworm *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 16653–16658.
- Sammeth, M. (2009). Complete Alternative Splicing Events Are Bubbles in Splicing Graphs. *J. Comput. Biol.* *16*, 1117–1140.
- Sánchez, L.V., and Ramírez, J.D. (2013). Congenital and oral transmission of American trypanosomiasis: an overview of physiopathogenic aspects. *Parasitology* *140*, 147–159.
- Sánchez-Gracia, A., and Rozas, J. (2008). Divergent evolution and molecular adaptation in the *Drosophila* odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. *BMC Evol. Biol.* *8*, 323.
- Sánchez-Gracia, A., Vieira, F.G., and Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity* *103*, 208–216.
- Sanchez-Martin, M.J., Feliciangeli, M.D., Campbell-Lendrum, D., and Davies, C.R. (2006a). Could the Chagas disease elimination programme in Venezuela be compromised by reinvasion of houses by sylvatic *Rhodnius prolixus* bug populations? *Trop. Med. Int. Health* *11*, 1585–1593.
- Sandler, B.H., Nikonova, L., Leal, W.S., and Clardy, J. (2000). Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol.* *7*, 143–151.
- Santos, A., Ribeiro, J.M.C., Lehane, M.J., Gontijo, N.F., Veloso, A.B., Sant'Anna, M.R.V., Nascimento Araujo, R., Grisard, E.C., and Pereira, M.H. (2007). The sialotranscriptome of the blood-sucking bug *Triatoma brasiliensis* (Hemiptera, Triatominae). *Insect Biochem. Mol. Biol.* *37*, 702–712.
- Sarov-Blat, L., So, W.V., Liu, L., and Rosbash, M. (2000). The *Drosophila* takeout Gene Is a Novel Molecular Link between Circadian Rhythms and Feeding Behavior. *Cell* *101*, 647–656.
- Sarquis, O., Sposina, R., Oliveira, T.G. de, Cord, M., Roberto, J., Cabello, P.H., Borges-Pereira, J., and Lima, M.M. (2006). Aspects of peridomiliary ecotopes in rural areas of Northeastern Brazil associated to triatomine (Hemiptera, Reduviidae) infestation, vectors of Chagas disease. *Mem. Inst. Oswaldo Cruz* *101*, 143–147.
- Sarquis, O., Carvalho-Costa, F.A., Oliveira, L.S., Duarte, R., D Andrea, P.S., de Oliveira, T.G., and Lima, M.M. (2010). Ecology of *Triatoma brasiliensis* in northeastern Brazil: seasonal distribution, feeding resources, and *Trypanosoma cruzi* infection in a sylvatic population. *J. Vector Ecol. J. Soc. Vector Ecol.* *35*, 385–394.
- Sarquis, O., Carvalho-Costa, F.A., Toma, H.K., Georg, I., Burgoa, M.R., and Lima, M.M. (2011). Eco-epidemiology of Chagas disease in northeastern Brazil: *Triatoma brasiliensis*, *T. pseudomaculata* and *Rhodnius nasutus* in the sylvatic, peridomestic and domestic environments. *Parasitol. Res.* *110*, 1481–1485.
- Scaloni, A., Monti, M., Angeli, S., and Pelosi, P. (1999). Structural analysis and disulfide-bridge pairing of two odorant-binding proteins from *Bombyx mori*. *Biochem. Biophys. Res. Commun.* *266*, 386–391.
- Schaafsma, E.J., Hapke, S.D., and Banfield, M.G. (2012). Bed Bug (*Cimex lectularius* L.) Population Composition as Determined by Baited Traps. *Insects* *3*, 442–452.
- Schmidt, K.A., and Ostfeld, R.S. (2001). Biodiversity and the dilution effect in disease ecology. *Ecology* *82*, 609–619.
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* *27*, 863–864.

- Schmunis, G.A. (2007). Epidemiology of Chagas disease in non endemic countries: the role of international migration. *Mem. Inst. Oswaldo Cruz* 102, 75–86.
- Schmunis, G.A., and Yadon, Z.E. (2010). Chagas disease: A Latin American health problem becoming a world health problem. *Acta Trop.* 115, 14–21.
- Schofield, C.J. (1979). The behaviour of Triatominae (Hemiptera: Reduviidae): a review. *Bull. Entomol. Res.* 69, 363–379.
- Schofield, C.J. (1980). Nutritional status of domestic populations of *Triatoma infestans*. *Trans. R. Soc. Trop. Med. Hyg.* 74, 770–778.
- Schofield, C.J., and Dujardin, J.P. (1999). Theories on the evolution of *Rhodnius*. *Actual Biol* 21, 183–197.
- Schofield, C.J., and Galvão, C. (2009). Classification, evolution, and species groups within the Triatominae. *Acta Trop.* 110, 88–100.
- Schofield, G.J., and Patterson, J.W. (1977). Assembly Pheromone of *Triatoma infestans* and *Rhodnius Prolixus* Nymphs (Hemiptera: Reduviidae). *J. Med. Entomol.* 13, 727–734.
- Schofield, C.J., Diotaiuti, L., and Dujardin, J.P. (1999). The process of domestication in triatominae. *Mem. Inst. Oswaldo Cruz* 94, 375–378.
- Schoonhoven, L.M., Loon, J.J.A. van, and Dicke, M. (2005). *Insect-plant biology*. xvii + 421 pp.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092.
- Scott, J.G. (1999). Cytochromes P450 and insecticide resistance. *Insect Biochem. Mol. Biol.* 29, 757–777.
- Sengupta, S., Bolin, J.M., Ruotti, V., Nguyen, B.K., Thomson, J.A., Elwell, A.L., and Stewart, R. (2011). Single Read and Paired End mRNA-Seq Illumina Libraries from 10 Nanograms Total RNA. *J. Vis. Exp.*
- Seyednasrollah, F., Laiho, A., and Elo, L.L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* 16, 59–70.
- Shaw, J., Lainson, R., and Fraiha, H. (1969). Observations on the epidemiology of the first autochthonous cases of Chagas' disease in Belém, State of Pará, Brazil. *Rev. Saúde Pública* 3, 153–157.
- Silveira, A.C., and Vinhaes, M.C. (1999). Elimination of vector-borne transmission of Chagas disease. *Mem. Inst. Oswaldo Cruz* 94, 405–411.
- Simon, J.-C., d' Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiéry, J., Nouhaud, P., Peccoud, J., Sugio, A., and Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Brief. Funct. Genomics* 14, 413–423.
- Slone, J., Daniels, J., and Amrein, H. (2007). Sugar Receptors in *Drosophila*. *Curr. Biol.* 17, 1809–1816.
- Smadja, C., Shi, P., Butlin, R.K., and Robertson, H.M. (2009). Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, *Acyrtosiphon pisum*. *Mol. Biol. Evol.* 26, 2073–2086.
- Snell-Rood, E.C., Van Dyken, J.D., Cruickshank, T., Wade, M.J., and Moczek, A.P. (2010). Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *BioEssays* 32, 71–81.
- So, W.V., Sarov-Blat, L., Kotarski, C.K., McDonald, M.J., Allada, R., and Rosbash, M. (2000). takeout, a Novel *Drosophila* Gene under Circadian Clock Transcriptional Regulation. *Mol. Cell. Biol.* 20, 6935–6944.
- Sombroek, W. (2001). Spatial and Temporal Patterns of Amazon Rainfall. *AMBIO J. Hum. Environ.* 30, 388–396.
- Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* 14, 91.
- Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. *Trends Ecol. Evol.* 25, 705–712.
- Sun, L., Zhou, J.-J., Gu, S.-H., Xiao, H.-J., Guo, Y.-Y., Liu, Z.-W., and Zhang, Y.-J. (2015). Chemosensillum immunolocalization and ligand specificity of chemosensory proteins in the alfalfa plant bug *Adelphocoris lineolatus* (Goeze). *Sci. Rep.* 5.
- Sun, Y.-L., Huang, L.-Q., Pelosi, P., and Wang, C.-Z. (2012). Expression in Antennae and Reproductive Organs Suggests a Dual Role of an Odorant-Binding Protein in Two Sibling *Helicoverpa* Species. *PLoS ONE* 7, e30040.
- Surget-Groba, Y., and Montoya-Burgos, J.I. (2010). Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res.* 20, 1432–1440.
- Tanaka, K., Uda, Y., Ono, Y., Nakagawa, T., Suwa, M., Yamaoka, R., and Touhara, K. (2009). Highly Selective Tuning of a Silkworm Olfactory Receptor to a Key Mulberry Leaf Volatile. *Curr. Biol.* 19, 881–890.

- Taneja, J., and Guerin, P.M. (1997). Ammonia attracts the haematophagous bug *Triatoma infestans*: behavioural and neurophysiological data on nymphs. *J. Comp. Physiol. A* *181*, 21–34.
- Teets, N.M., Peyton, J.T., Colinet, H., Renault, D., Kelley, J.L., Kawarasaki, Y., Lee, R.E., and Denlinger, D.L. (2012). Gene expression changes governing extreme dehydration tolerance in an Antarctic insect. *Proc. Natl. Acad. Sci.* *109*, 20744–20749.
- Teixeira, A.R., Monteiro, P.S., Rebelo, J.M., Argañaraz, E.R., Vieira, D., Lauria-Pires, L., Nascimento, R., Vexenat, C.A., Silva, A.R., Ault, S.K., et al. (2001). Emerging Chagas disease: trophic network and cycle of transmission of *Trypanosoma cruzi* from palm trees in the Amazon. *Emerg. Infect. Dis.* *7*, 100–112.
- Teixeira, A.R.L., Nascimento, R.J., and Sturm, N.R. (2006). Evolution and pathology in Chagas disease - A review. *Mem. Inst. Oswaldo Cruz* *101*, 463–491.
- Thibert-Plante, X., and Hendry, A.P. (2011). The consequences of phenotypic plasticity for ecological speciation. *J. Evol. Biol.* *24*, 326–342.
- Thomson, M.C. (2014). Emerging Infectious Diseases, Vector-Borne Diseases, and Climate Change. In *Global Environmental Change*, B. Freedman, ed. (Springer Netherlands), pp. 623–628.
- Togawa, T., Dunn, W.A., Emmons, A.C., Nagao, J., and Willis, J.H. (2008). Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* *38*, 508–519.
- Tomaselli, S., Crescenzi, O., Sanfelice, D., AB, E., Wechselberger, R., Angeli, S., Scaloni, A., Boelens, R., Tancredi, T., Pelosi, P., et al. (2006). Solution Structure of a Chemosensory Protein from the Desert Locust *Schistocerca gregaria*. *Biochemistry (Mosc.)* *45*, 10606–10613.
- Toso M, A., Vial U, F., and Galanti, N. (2011). Transmisión de la enfermedad de Chagas por vía oral. *Rev. Médica Chile* *139*, 258–266.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* *28*, 511–515.
- Valente, S.A. da S., Valente, V. da C., and Fraiha Neto, H. (1999). Considerations on the epidemiology and transmission of Chagas disease in the Brazilian amazon. *Mem. Inst. Oswaldo Cruz* *94*, 395–398.
- Vazquez-Prokopec, G.M., Ceballos, L.A., Kitron, U., and Gürtler, R.E. (2004). Active Dispersal of Natural Populations of *Triatoma infestans* (Hemiptera: Reduviidae) in Rural Northwestern Argentina. *J. Med. Entomol.* *41*, 614–621.
- Vieira, F.G., and Rozas, J. (2011). Comparative Genomics of the Odorant-Binding and Chemosensory Protein Gene Families across the Arthropoda: Origin and Evolutionary History of the Chemosensory System. *Genome Biol. Evol.* *3*, 476–490.
- Vieira, F.G., Sánchez-Gracia, A., and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol.* *8*, R235.
- Vienne, D.M. de, Ollier, S., and Aguileta, G. (2012). Phylo-MCOA: A Fast and Efficient Method to Detect Outlier Genes and Species in Phylogenomics Using Multiple Co-inertia Analysis. *Mol. Biol. Evol.* msr317.
- Vijay, N., Poelstra, J.W., Künstner, A., and Wolf, J.B.W. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol. Ecol.* *22*, 620–634.
- Vitta, A.C.R., and Lorenzo, M.G. (2009). Copulation and Mate Guarding Behavior in *Triatoma brasiliensis* (Hemiptera: Reduviidae). *J. Med. Entomol.* *46*, 789–795.
- Vitta, A.C.R., Bohman, B., Unelius, C.R., and Lorenzo, M.G. (2009). Behavioral and Electrophysiological Responses of *Triatoma brasiliensis* Males to Volatiles Produced in the Metasternal Glands of Females. *J. Chem. Ecol.* *35*, 1212–1221.
- Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet.* *47*, 97–120.
- Vogt, R.G., and Riddiford, L.M. (1981). Pheromone binding and inactivation by moth antennae. *Nature* *293*, 161–163.
- Vogt, R.G., Prestwich, G.D., and Lerner, M.R. (1991). Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *J. Neurobiol.* *22*, 74–84.

- Vogt, R.G., Rogers, M.E., Franco, M., and Sun, M. (2002). A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in *Manduca sexta* (Lepidoptera) and the organization of OBP genes in *Drosophila melanogaster* (Diptera). *J. Exp. Biol.* *205*, 719–744.
- Vontas, J., Small, G., and HEMINGWAY, J. (2001). Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. *Biochem J* *357*, 65–72.
- Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A., and Axel, R. (1999). A Spatial Map of Olfactory Receptor Expression in the *Drosophila* Antenna. *Cell* *96*, 725–736.
- Walter, A., Rego, I.P. do, Ferreira, A.J., and Rogier, C. (2005). Risk factors for reinvasion of human dwellings by sylvatic triatomines in northern Bahia State, Brazil. *Cad. Saúde Pública* *21*, 974–978.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* *10*, 57–63.
- Wanner, K.W., Willis, L.G., Theilmann, D.A., Isman, M.B., Feng, Q., and Plettner, E. (2004). Analysis of the insect os-d-like gene family. *J. Chem. Ecol.* *30*, 889–911.
- Ward, J.P. (1981). A comparison of the behavioural responses of the haematophagous bug, *Triatoma infestans*, to synthetic homologues of two naturally occurring chemicals (n- and iso-butyric acid). *Physiol. Entomol.* *6*, 325–329.
- Ward, J.A., Ponnala, L., and Weber, C.A. (2012). Strategies for transcriptome analysis in nonmodel plants. *Am. J. Bot.* *99*, 267–276.
- Watanabe, K., Toba, G., Koganezawa, M., and Yamamoto, D. (2011). Gr39a, a Highly Diversified Gustatory Receptor in *Drosophila*, has a Role in Sexual Behavior. *Behav. Genet.* *41*, 746–753.
- Weeks, E.N.I., Birkett, M.A., Cameron, M.M., Pickett, J.A., and Logan, J.G. (2011). Semiochemicals of the common bed bug, *Cimex lectularius* L. (Hemiptera: Cimicidae), and their potential for use in monitoring and control. *Pest Manag. Sci.* *67*, 10–20.
- Weiss, L.A., Dahanukar, A., Kwon, J.Y., Banerjee, D., and Carlson, J.R. (2011). The Molecular and Cellular Basis of Bitter Taste in *Drosophila*. *Neuron* *69*, 258–272.
- Werner, T. (2010). Next generation sequencing in functional genomics. *Brief. Bioinform.* *11*, 499–511.
- Wigglesworth, V.B. (1936). Memoirs: The function of the Corpus Allatum in the Growth and Reproduction of *Rhodnius Prolixus* (Hemiptera). *Q. J. Microsc. Sci.* *s2-79*, 91–121.
- Wilhelm, B.T., Marguerat, S., Goodhead, I., and Bähler, J. (2010). Defining transcribed regions using RNA-seq. *Nat. Protoc.* *5*, 255–266.
- Wilson, R.I., and Mainen, Z.F. (2006). Early Events in Olfactory Processing. *Annu. Rev. Neurosci.* *29*, 163–201.
- Witzgall, P., Kirsch, P., and Cork, A. (2010). Sex Pheromones and Their Impact on Pest Management. *J. Chem. Ecol.* *36*, 80–100.
- World Health Organization (2015). Weekly epidemiological record. *90*, 33–44.
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., et al. (2013). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *ArXiv13056760 Q-Bio*.
- Xu, L., Wu, M., and Han, Z. (2013). Overexpression of Multiple Detoxification Genes in Deltamethrin Resistant *Laodelphax striatellus* (Hemiptera: Delphacidae) in China. *PLoS ONE* *8*, e79443.
- Xu, P., Atkinson, R., Jones, D.N.M., and Smith, D.P. (2005). *Drosophila* OBP LUSH Is Required for Activity of Pheromone-Sensitive Neurons. *Neuron* *45*, 193–200.
- Xuan, N., Guo, X., Xie, H.-Y., Lou, Q.-N., Lu, X.-B., Liu, G.-X., and Picimbon, J.-F. (2015). Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. *Insect Sci.* *22*, 203–219.
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* *24*, 1586–1591.
- Zacharias, C.A., Pontes, G.B., Lorenzo, M.G., and Manrique, G. (2010). Flight Initiation by Male *Rhodnius prolixus* is Promoted by Female Odors. *J. Chem. Ecol.* *36*, 449–451.
- Zhang, J., Goyer, C., and Pelletier, Y. (2008). Environmental stresses induce the expression of putative glycine-rich insect cuticular protein genes in adult *Leptinotarsa decemlineata* (Say). *Insect Mol. Biol.* *17*, 209–216.
- Zhang, J., Feng, J., Lu, J., Yang, Y., Zhang, X., Wan, D., and Liu, J. (2014a). Transcriptome differences between two sister desert poplar species under salt stress. *BMC Genomics* *15*, 337.
- Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., et al. (2014b). A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. *PLoS ONE* *9*, e103207.
- Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X., and Hao, P. (2011). Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* *12*, S2.

- Zhou, J.-J., Robertson, G., He, X., Dufour, S., Hooper, A.M., Pickett, J.A., Keep, N.H., and Field, L.M. (2009). Characterisation of *Bombyx mori* Odorant-binding Proteins Reveals that a General Odorant-binding Protein Discriminates Between Sex Pheromone Components. *J. Mol. Biol.* *389*, 529–545.
- Zhou, J.-J., Vieira, F.G., He, X.-L., Smadja, C., Liu, R., Rozas, J., and Field, L.M. (2010a). Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrthosiphon pisum*. *Insect Mol. Biol.* *19*, 113–122.
- Zhou, J.-J., Field, L.M., and He, X.L. (2010b). Insect Odorant-Binding Proteins: Do They Offer an Alternative Pest Control Strategy? *Outlooks Pest Manag.* *21*, 31–34.
- Zhou, X., Slone, J.D., Rokas, A., Berger, S.L., Liebig, J., Ray, A., Reinberg, D., and Zwiebel, L.J. (2012). Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding. *PLoS Genet* *8*, e1002930.
- Zhu, F., Gujar, H., Gordon, J.R., Haynes, K.F., Potter, M.F., and Palli, S.R. (2013). Bed bugs evolved unique adaptive strategy to resist pyrethroid insecticides. *Sci. Rep.* *3*, 1456.
- Zhu, H., Casselman, A., and Reppert, S.M. (2008). Chasing Migration Genes: A Brain Expressed Sequence Tag Resource for Summer and Migratory Monarch Butterflies (*Danaus plexippus*). *PLoS ONE* *3*, e1345.
- Zingales, B., Andrade, S.G., Briones, M.R.S., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., et al. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem. Inst. Oswaldo Cruz* *104*, 1051–1054.

Titre : Le processus de domiciliation des punaises hématophages vectrices de la maladie de Chagas : apport de l'étude du transcriptome chimiosensoriel

Mots clés : Adaptation, Anthroposystèmes, Domiciliation, Maladie de Chagas, RNAseq, Système chimiosensoriels

Résumé : En Amérique Latine, les punaises hématophages Triatominae transmettent à l'homme le parasite *Trypanosoma cruzi*, responsable de la maladie de Chagas. Même si les programmes d'éradication chimique des vecteurs sont efficaces, la maladie persiste du fait de la recolonisation des habitations humaines par des vecteurs provenant d'habitats naturels. Ainsi, certaines espèces présentent une capacité d'adaptation aux anthroposystèmes (processus de domiciliation), alors que d'autres espèces apparentées ne l'ont pas. Comprendre cette capacité d'adaptation est crucial d'un point de vue épidémiologique afin de cibler les espèces présentant un risque pour l'homme. La capacité à s'adapter à un nouvel habitat pourrait être liée à l'évolution du répertoire de gènes du système chimiosensoriel, important pour la perception du milieu.

Cette étude a porté sur le système chimiosensoriel des Triatominae dans le but de documenter le processus d'adaptation et donc de domiciliation des vecteurs. Des données transcriptomiques obtenues en séquençage à haut débit ont été utilisées pour annoter et répertorier les gènes chimiosensoriels ainsi que pour comparer leur expression au sein de punaises hématophages d'habitats différents.

Différentes méthodes ont été utilisées pour assembler les transcriptomes de référence de plusieurs espèces de Triatominae.

Chez l'ensemble des espèces de Triatominae étudiées, l'analyse du répertoire des gènes chimiosensoriels codant les OBPs et CSPs comparé à celui d'autres Paraneoptères a montré des expansions géniques pouvant refléter des processus adaptatifs.

Les analyses d'expression différentielle concernant les différentes populations de *T. brasiliensis* et les espèces *R. prolixus*/*R. robustus* ont montré qu'un certain nombre de transcrits sont différentiellement exprimés selon l'environnement dans lequel ont évolué les punaises notamment des gènes chimiosensoriels (OBPs, CSPs) ainsi que des gènes impliqués dans le rythme circadien et le comportement de recherche alimentaire (*takeout*) et dans la réponse à des stress environnementaux.

Ce travail a permis de mettre à la disposition de la communauté scientifique des outils performants pour l'étude du processus de domiciliation des vecteurs de la maladie de Chagas (transcriptome, répertoire de gènes). Il a également permis de révéler des gènes qui pourraient être impliqués dans l'adaptation et/ou la plasticité phénotypique en réponse à un changement d'habitat. La compréhension des bases moléculaires de l'adaptation des vecteurs aux habitations humaines ouvre des potentialités de développer des méthodes alternatives de lutte contre les vecteurs qui pourraient être basées sur une perturbation de la communication chimique.

Title : Domiciliation process of bloodsucking bugs, vectors of Chagas disease : contribution of the chemosensory transcriptome study

Keywords : Adaptation, Anthroposystems, Domiciliation, Chagas disease, RNAseq, Chemosensory system

Abstract : In Latin America, the bloodsucking bugs are vectors of the parasite *Trypanosoma cruzi*, which causes Chagas disease. Even if chemical control campaigns are effective against vectors, the disease persists due to the recolonization of human habitations by vectors from natural habitats. Some species have the capacity to adapt to anthroposystems (domiciliation process), while other related species do not. Understanding this capacity to adapt is crucial from an epidemiological perspective to target species at risk to humans. The capacity to adapt to a new habitat could be linked to changes in the repertoire of chemosensory system genes, important to detect various odor stimuli.

This study is based on the chemosensory system of Triatominae to document the adaptation process and then the domiciliation of the vectors. Transcriptomic data obtained by high-throughput sequencing were used to annotate and list the chemosensory genes and also to compare their expression in bloodsucking bugs from different habitats.

Various methods have been used to assemble the reference transcriptomes of several Triatominae species.

Among all studied Triatominae species, the chemosensory gene repertoire showed a high diversity and genic expansions compared to that of others Paraneoptera, which could reflect adaptive process.

The differential expression analyses on the *T. brasiliensis* populations and the *R. prolixus* / *R. robustus* species showed that some transcripts are differentially expressed according to the environment in which the bugs have evolved, especially the chemosensory genes (OBP, CSP) and also genes involved in the circadian rhythm and foraging behavior (*takeout*), and in the response to environmental stress.

This work has helped make available to the scientific community powerful tools for studying the process of domiciliation of Chagas disease vectors (transcriptome, gene repertoire). It also revealed genes that could be involved in the adaptation and/or phenotypic plasticity in response to a change in habitat.

Understanding the molecular basis of vector adaptation to human dwellings opens the potential to develop new tools to control the disease vectors, for example by disrupting chemical communication.