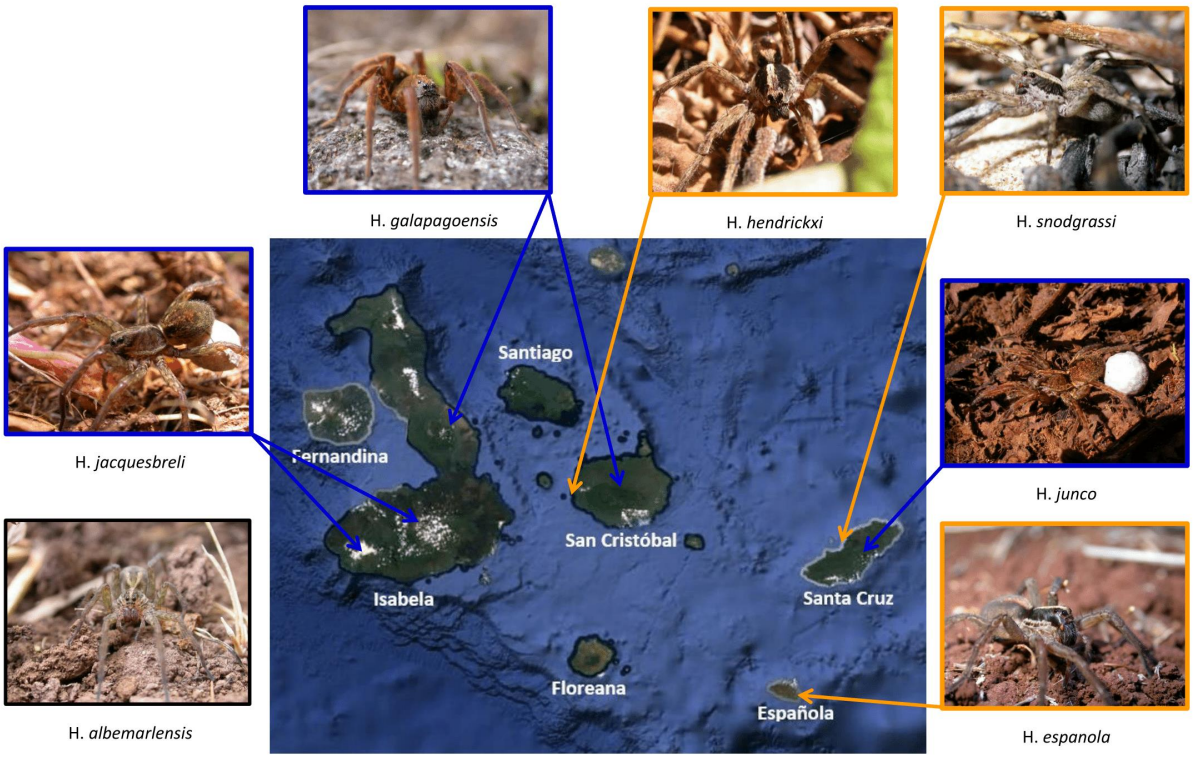
Faculty of Sciences

Researchgroup: Terrestrial Ecology Unit (TEREC)

Academic year 2015 / 2016

Population genomic analysis of a parallel wolf spider radiation from the Galápagos

De Corte Zoë



Supervisor: Prof. Dr. Frederik Hendrickx

Co-supervisor: Prof. Dr. Carl Vangestel

Thesis submitted to obtain the degree of Master of Science in Biology

Content

1. Acknowledgments

My most sincere gratitude and appreciation goes out to Prof. Dr. Frederik Hendrickx and Prof. Dr. Carl Vangestel for giving me the opportunity to participate in such an incredibly interesting research and giving me the opportunity to learn to use these new techniques. You have provided me with help and guidance, from making the RAD libraries, processing the obtained reads, and analysing the NGS data. I learned a great deal from the interesting discussions and the great collaboration we had. Thank you for your useful comments and remarks which improved my thesis tremendously.

I also would like to thank Léon Baert, Charlotte De Busschere, Wouter Dekoninck, Steven Van Belleghem, Frederik Hendrickx & Carl Vangestel for collecting the samples in the Galapagos. I would, of course, love to return the favour sometime in the future. Katrien, thank you for taking one of my RAD library samples to Leuven for the shearing step.

Furthermore, I thank my colleagues from the KBIN for the nice working environment and relaxing lunch breaks. You made working in the KBIN a truly wonderful experience.

I want to express my very great appreciation to Roeland for taking the time to read and correct my thesis.

And last but not least my honest thanks to my family and friends for their patience and support during the whole process.

2. Introduction

As different islands often harbor similar environmental gradients, island radiations sometimes result in a most intriguing phenomenon, where phenotypically similar species, i.e. ecotypes, evolve recurrently on each distinct island (Gillespie 2013). As it is highly improbable that the occurrence of these repeated phenotypes is the product of chance, they can therefore be interpreted as evidence that natural selection causes adaptation to new environments (Schluter 2000). Notwithstanding the great scientific interest this has generated among evolutionary biologists, the question whether these ecotypes arise from different genotypic changes or from shared ancestry remains largely unanswered to this day (Arendt and Reznick 2008; Pascoal et al. 2014; Rosenblum, Parent, and Brandt 2014). Island groups are the perfect natural laboratories to infer patterns of evolution as they are often small in size, have distinct boundaries and are geographically isolated (Losos and Ricklefs 2009).

The evolution of different ecotypes on island groups can occur in sympatry or allopatry (Losos and Ricklefs 2009). The two scenarios start with the same situation, where an ancestral species has colonized an island in the archipelago and colonizes the other islands afterwards. In allopatry, the populations on different islands become genetically differentiated and evolve to become different species. These species can subsequently disperse and colonize the other islands in which they will settle in comparable habitats, i.e. a 'species sorting ' mechanism (Leibold et al. 2004). In sympatry, the populations on the same island diverge and adapt to different niches or habitats (De Busschere et al. 2010). This is referred to as convergent or parallel evolution. Given the recent debate on the distinction between convergence and parallelism in evolution (Arendt and Reznick 2008; Gompel and Prud'homme 2009; Rosenblum, Parent, and Brandt 2014; Stern 2013), we refer to evolution of similar phenotypes, irrespective of the underlying mechanism, as parallel evolution .

Parallel evolution can emerge in several ways. The main distinction between the different alternatives is whether a different or identical genetic background leads to the same phenotype. The first case has been shown in a study of the evolution of flat wings in a Hawaiian cricket radiation, which revealed that different genomic regions are linked to the development of similar wing morphologies on different islands (Pascoal et al. 2014). Parallel evolution starting from the same genetic background can arise by three mechanisms. First, the same mutation that leads to a certain phenotype has occurred independently in different populations. Second, by the repeated selection of the same allele that was present in the ancestral population as standing genetic variation (Barrett and Schluter 2008). This mechanism has been proposed in the repeated evolution of the lateral plate reduction in freshwater populations out of marine populations. Parallel evolution by standing genetic variation is presumed to be easier and more common compared to new mutations, as the genetic variation is immediately available, and because it has already been passed through a selective filter and is initially present in higher frequencies. The origin of beneficial mutations in contrast are expected to be rare events, and are initially extremely rare in populations, which make them very prone to disappearing through genetic drift (Barrett and Schluter 2008). Third, adaptive alleles may be introduced in populations or species from other populations that evolved under the same selective pressures. This process is often referred to as adaptive gene introgression (Stern 2013). This mechanism is only recently being recognized in evolutionary research, even though the actual rate of occurrence of this phenomenon is one of the most debated topics in evolutionary biology (Garant, Forde, and Hendry 2007; Nosil 2008; Smadja and Butlin 2011). Gene flow is generally seen as a process that constrains adaptive divergence due to a decrease in genetic variation between populations and an increase within populations (Futuyma 2013; Garant, Forde, and Hendry 2007). However, the increase in genetic variation could increase the adaptive potential (Garant, Forde, and Hendry 2007; Smadja and Butlin 2011) and release constraints that are caused

by genetic correlations (Seehausen et al. 2014). The process of adaptive gene introgression might even be accelerated if individuals do not disperse at random, but settle in the habitat that best matches their phenotype ('matching habitat choice ' cfr. Edelaar, Siepielski, and Clobert 2008). Divergence can occur in the face of gene flow as illustrated by case-studies on *Heliconius* butterflies (Martin et al. 2013), Darwin's finches (Lamichhaney et al. 2015), Tennessee cave salamanders (Niemiller, Fitzpatrick, and Miller 2008) and the rough periwinkle, *Littorina saxatilis* (Butlin et al. 2014).

Genomics can help us to determine the mode of speciation and the underlying mechanism (Martin et al. 2013; Seehausen et al. 2014) as advances of next-generation sequencing (NGS) enable us to uncover thousands of markers across a genome (Davey et al. 2011). Here, we make use of such technology in an attempt to unravel the evolutionary history of a parallel radiation of wolf spiders of the genus *Hogna* Simon, 1885 (Lycosidae) at the Galápagos. Within this radiation, two different ecotypes repetitively co-occur on different islands, rendering this ecosystem a natural evolutionary experiment and a perfect candidate to study parallel speciation (Losos and Ricklefs 2009; Parent, Caccone, and Petren 2008).

The two different ecotypes can be identified based on their morphology (Figure 1), behaviour and the ecological zone they are occupying. The first group, further referred to as 'top species ' live at high elevations on the islands in the dense pampa vegetation. This group consists of H. *galapagoensis* (Banks 1902) on Santa Cruz, Santiago and Isabela (Volcan Cerro Azul and Volcan Alcedo), H. *jacquesbreli* Baert & Maelfait, 2008 on Isabela (Volcan Sierra Negra and Volcan Cerro Azul) and H. *junco* Baert & Maelfait, 2008 on San Cristóbal (Baert, Maelfait, and Hendrickx 2008). The second group consists of species that live in the dry supralittoral and arid zone along the coast, in the vegetated dunes and open shrub land, and will be referred to as 'low' species. The group consists of H. *snodgrassi* (Banks 1902) on San Cristóbal, H. *espanola* Baert & Maelfait, 2008 on Española and H. *hendrickxi* Baert & Maelfait, 2008 on Santa Cruz (Baert, Maelfait, and Hendrickx 2008). Beside these two different ecotypes, a more generalistic species, H. *albemarlensis* (Banks 1902) occurs in a wide range of habitats with high humidity. H. *albemarlensis* might be the result of a separate colonization, as this species is widespread, occupies a wide variety of habitats and shows a strong reduction in genetic variation compared to the other species (De Busschere et al. 2010). The *Hogna* species occupies the islands Española, San Cristóbal , Santa Cruz, Santiago, Isabela. Isabela, Santiago and San Cristóbal belong to the 'core' islands, which are located within the 200m isobaths from each other. Since 700ka, the 'core' islands have experienced major changes to distribution of land and sea due to fluctuating sea levels. Lower sea levels probably caused higher connectivity between the populations on different islands.

A majority of studies on parallel speciation have focused on traits influenced by natural selection. However, these conclusions may strongly differ from those based on traits subjected to sexual selection. A previous study using a large set of genital and non-genital traits demonstrated two different morphological divergence patterns within the *Hogna* radiation (De Busschere et al. 2012). Variations in colour and biometrics strongly covaried with ecotypic differentiation, while divergence in genital traits did not adhere to such an evolutionary scenario, but rather reflected phylogenetic relationships and showed high resemblance of genital traits between different ecotypes within a single island for some instances (Figure 2).

We will attempt to infer which mechanisms underlie the repeated evolution of the same ecotypes within island groups. The wolf spider genus *Hogna* Simon, 1885 (Lycosidae) is perfect for addressing this question, as it consists of both inter-island and intra-island sister-species pairs (Warren et al. 2015). To investigate this, we will determine the population structure, patterns of genomic differentiation and the phylogenetic relationship based on genome-wide data.

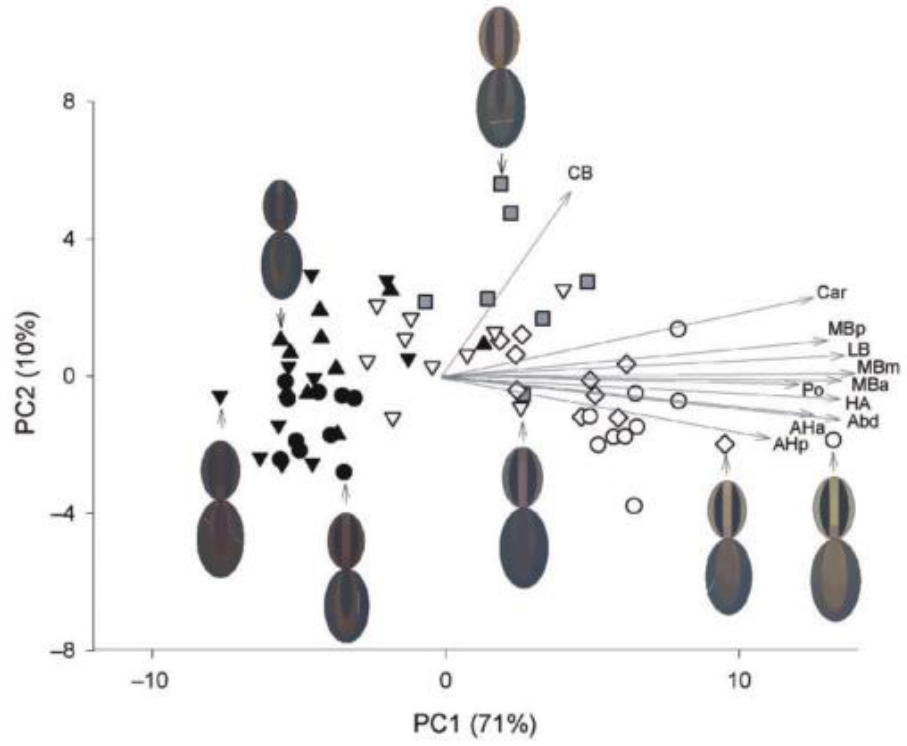


Figure 1-Colour variation of female individuals of *Hogna* species from Galápagos depicted in a PCA ordination, with high-elevation species depicted in black [H. *junco* (circles); H. *galapagoensis* (inverted triangles); H. *jacquesbreli* (upright triangles)] and low-elevation species depicted with open symbols [H. *snodgrassi* (circles); H. *hendrickxi* (inverted tri- angles); H. *espanola* (diamonds)] and H. *albemarlensis* (grey squares). The gradient in overall darkness is visualized by means of schematic pictures based on the colour composition of the measured traits. Eigenvectors are scaled by the eigenvalues of the respective PC axis. Measured traits are carapace (Car), abdomen (Abd), lateral band (LB), carapace band (CB), anterior & posterior region around heart spot (Aha & AHp), heart spot (HA), anterior, median, posterior part of median band (MBa, MBm & MBp) and posterior part of abdomen (Po)(De Busschere et al. 2012, figure 3).

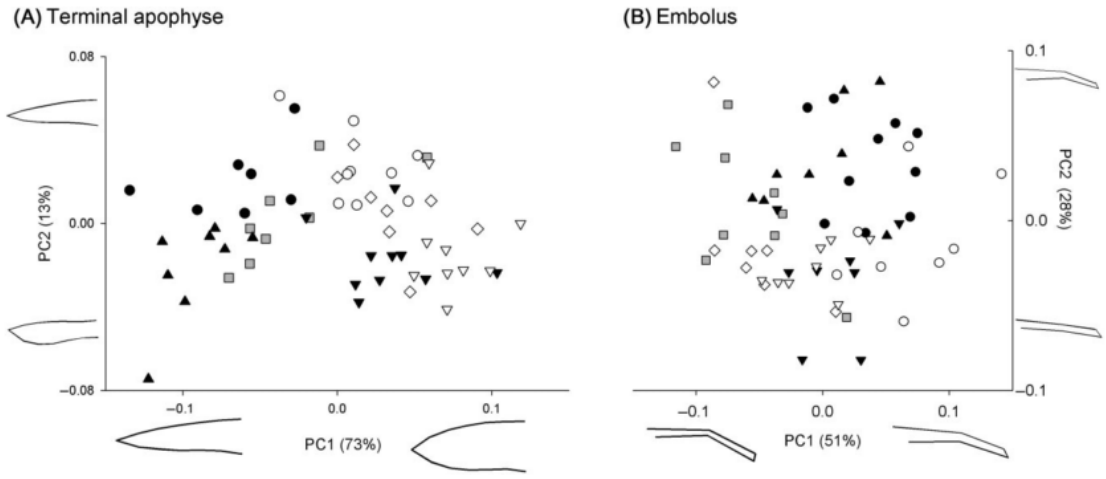


Figure 2-Shape differences in two male genital structures: (A) terminal apophysis and (B) embolus with high-elevation species depicted in black [H. *junco* (circles); H. *galapagoensis* (inverted triangles); H. *jacquesbreli* (upright triangles)] and low-elevation species depicted with open symbols [H. *snodgrassi* (circles); H. *hendrickxi* (inverted triangles); H. *espanola* (diamonds)] and H. *albemarlensis* (grey square).(Charlotte De Busschere et al. 2012, figure 5)

3. Objectives

Here, we aim to investigate which mechanisms underlie the repeated evolution of similar ecotypes of a *Hogna* radiation on the Galápagos islands. To this end, we will explore patterns of genetic diversity by determining (i) the population structure, (ii) regions of strong genomic divergence and (iii) phylogenetic relationships based on genome-wide data.

Population structure

By applying various clustering algorithms, we will attempt to delineate the optimal number of genetic clusters. In addition, we will quantify whether genetic variation will group predominantly according to (i) geography or (ii) ecotype. Grouping according to geography indicates within island radiation (Losos and Ricklefs 2009) while grouping according to ecotype indicates the evolution of the ecotype occurred once and subsequently dispersed within the archipelago (Pascoal et al. 2014). De Busschere et al. 2012 have shown, for a limited amount of markers, that the 'high elevation' ecotype and 'coastal dry' ecotype evolved within islands and we hypothesize that the same pattern will be observed for the genome-wide data.

Patterns of genomic differentiation

In order to separate neutral from adaptive genetic variation, we will perform a genome-wide scan for signatures of selection on *Hogna* populations of Santa Cruz and San Cristóbal. Both ecotypes inhabit these islands, allowing us to estimate levels of genetic differentiation between i) different ecotypes within an island and ii) similar ecotypes residing on different islands. Statistical outliers, i.e. genetic variants characterized by an unusually strong genetic divergence, are putative candidate loci involved in local adaptation. Therefore, the aim of this genome-wide screening is twofold. Firstly, we will assess to what extent loci of strong between-ecotype divergence are shared among both islands, indicating similar genetic and developmental pathways underlying the repeated evolution of ecotypes (Arendt and Reznick 2008). Secondly, we will explore to what extent island-specific outliers can be detected, reflecting geographical or reproductive isolation as a key agent in shaping the spatial distribution of genetic variation (De Busschere et al. 2012).

Phylogenetic relationship

We aim to explore to what extent the recurrent appearance of similar ecotypes on each island is the result of (i) independent parallel evolution or (ii) rather the outcome of introgressive hybridization, where a single speciation event and subsequent spread of adaptive variation to other islands facilitated speciation. Both aforementioned scenarios correspond to different evolutionary pathways, and are consequently characterized by distinct genomic signatures. A key component of the analytical workflow will entail contrasting phylogenies of neutral and adaptive genomic regions. Under parallel speciation, we expect similar phylogenetic topologies for both neutral and adaptive genomic regions. Also, haplotypes will tend to cluster according to geography (De Busschere et al. 2010). In contrast, when convergent phenotypes have evolved through a process such as hybridization, adaptive and neutral genes do not share the same evolutionary history, and hence will result in a phylogenetic incongruence. Neutral genes show a phylogenetic pattern identical to that of parallel speciation, while haplotypes of adaptive genes will cluster according to ecotype.

4. Material & methods

4.1. Sampling

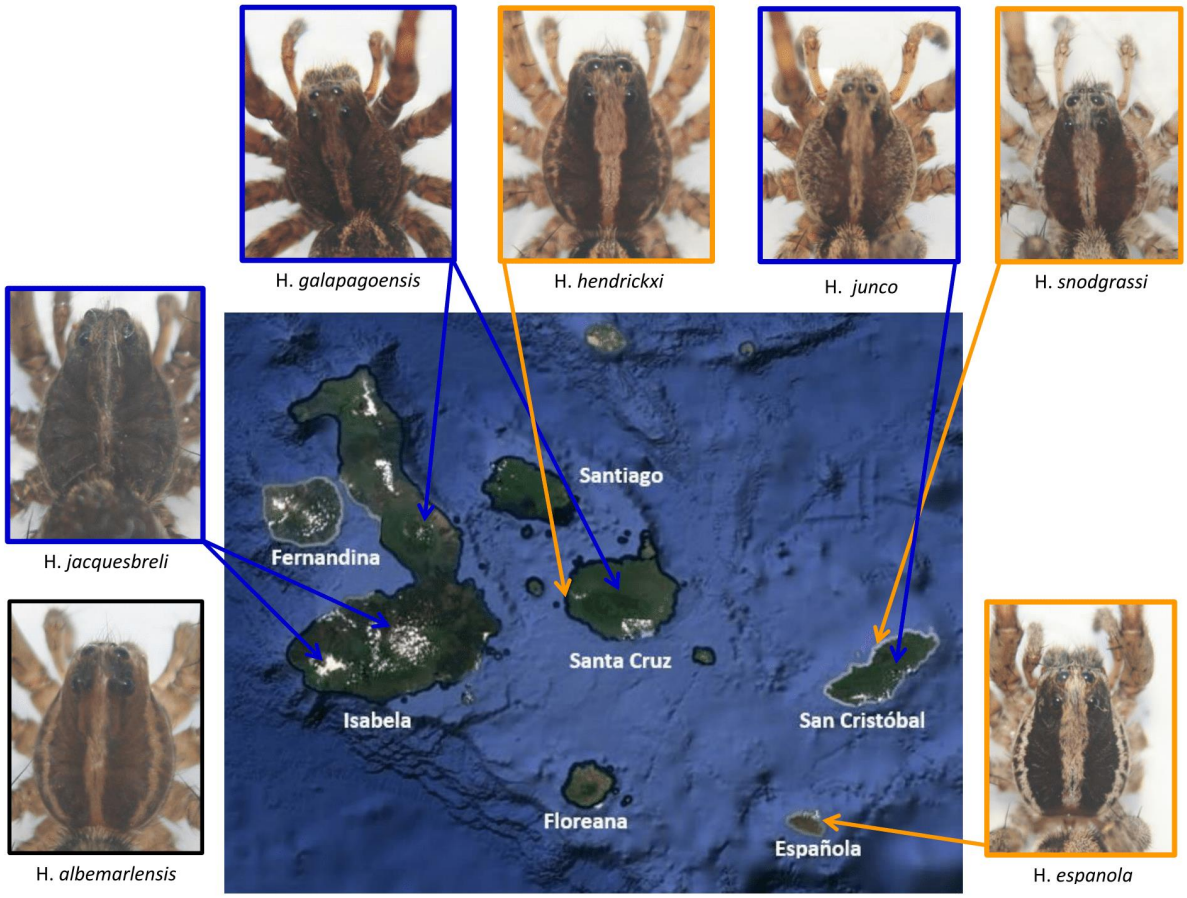


Figure 3. Geographical distribution of the genus *Hogna* on Galápagos with high elevation species (blue frame) and coastal dry species (yellow frame).

Samples were collected in 2009 by Léon Baert, Charlotte De Busschere, Wouter Dekoninck & Frederik Hendrickx and in 2014 by Wouter Dekoninck, Frederik Hendrickx, Steven Van Belleghem & Carl Vangestel (Figure 3). All specimens were captured alive and a leg was immediately removed and transferred in 97% ethanol. Morphometric analysis has been conducted on these samples and the results are published in the *Biological Journal of the Linnean Society* (De Busschere et al. 2012)

4.2. DNA extraction, RAD library preparation and genome assembly

DNA extraction

DNA was extracted from the legs of 96 individuals (Appendix), using the NucleoSpin Tissue kit (Macherey-Nagel GmBH) following the manufacturer's instructions. The final DNA concentration of 12 individuals was too low and was excluded in further preparations.

RAD library preparation

To obtain genome wide information on allele frequencies, we made use of Restriction-site Associated DNA sampling (RADseq) (Davey and Blaxter 2010). The RAD libraries were prepared according to the protocol described in Etter et al. (2011) and (Baird et al. 2008). The individual DNA samples are digested using the

restriction enzyme SbfI-HF (NEB), which has an 8bp recognition site and cutsite (5'-CCTGCA^GG-3'). Individually barcoded P1 adapters are ligated to the fragment's overhanging end. The uniquely barcoded samples are pooled in multiplex RAD libraries (Library 71-76, appendix), consisting of 16 individuals each. The RAD libraries are sheared to an optimal size of 350bp and the fragments between 200-700bp are selected by gel size selection. The DNA fragments in the RAD libraries are blunted, followed by A-tailing and P2 adapter ligation. During enrichment PCR, a second uniquely barcoded adapter (P2) is ligated to the DNA fragments per library. The unique P1 and P2 adapters enable us to identify the samples. The sequencing will be performed on two Illumina lanes. Six libraries, including in total 84 individuals (appendix), were sequenced. Five libraries (Library 71-73, 75-76) were sequenced on an Illumina HiSeq1500 platform at the Medical Genetics institute of the UAntwerpen. This platform generates 100 bp paired-end . In order to obtain longer sequences of the different species, which are more useful for haplotype reconstruction and *de novo* RADtag assembly, one library (Library 74) was sequenced on Illumina MiSeq platform at the same institute. This sequencing resulted 300bp paired-end sequences.

Demultiplex and filter

The Stacks software package (Catchen et al. 2011; Catchen et al. 2013) was used to demultiplex the reads from each library, to eliminate low quality reads and optimize read number from the Illumina sequencing run. First, the process_radtags utility was used to determine if the barcode and the RAD cut sites of the raw reads are intact and subsequently demultiplexed. The average quality score is checked by the use of a sliding window down the length of the read. The length of this sliding window is by default 15% of the length of the read. Reads with scores that drop below 90% probability of being correct (a raw phred score of 10) were discarded.

The removed reads of lane 1 were much higher compared to the removed reads in lane 2. Quality of the reads per library and lane were assessed with the software package FastQC (Andrew 2010). As the quality at the ends of the reverse reads of lane 2 were low, resulting in a significant loss of the paired reads after the process_radtags utility, reverse reads of lane 1 were trimmed using the software FastXToolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Several trimming lengths are tested to assess the optimum trimming length to obtain the highest number of reads with removing the least amount of bp.

Next, PCR duplicates were removed with the clone_filter module in Stacks. This module identifies sequences with an identical forward and reverse read, and subsequently retains only one of those in the output file. As clone_filter cannot process reads of different lengths, reads of lane 1 and 2 were concatenated after the removal of PCR duplicates.

The reads of the MiSeq had different lengths, which made it impossible to remove PCR duplicates.

Assembly reference tags & mapping

Extracting genotype information for each tag and individual was performed by first constructing a set of reference sequences for each RADtag, followed by mapping of reads to this reference set and SNP calling by means of the GATK pipeline.

The set of reference RADtags was assembled based on the MiSeq reads of one H. *galapagoensis* male (H409_008) from Santa Cruz, this species was chosen as it belongs to the species group which we will further focus on. The Software Velvet v1.2.10 (Zerbino and Birney 2008) is used, it assembles the sequences by means of a De Bruijn graph and subsequently uses a set of algorithms to manipulate the de Bruijn graphs

to eliminate errors and resolve repeats. A kmer length of 31 was used for the assembly, average insert length was set to 550bp and the readcategory was 'shortPaired'.

The reads of each individual were aligned against the H. *galapagoensis* reference RADtags with Stampy v.1.0.28 using default settings (Lunter and Goodson 2011). A genome file and hash file is built and paired-end mapping was done.

The UnifiedGenotyper from the Genome Analysis Toolkit (GATK) (DePristo et al. 2011; McKenna et al. 2010) was used to call SNPs and to estimate the most likely genotypes of each individual. Also non-variable sites were called and outputted in the final Variance Call Format (VCF) file.

Phasing

Haplotypes were reconstructed by means of both Readbacked phasing, which uses read information to phase SNPs, and a population based maximum likelihood framework. Readbacked phasing was performed with the Readbackphasing tool from GATK (DePristo et al. 2011; McKenna et al. 2010) with a genotype quality threshold of 20. Only those SNPs that are contained within a single read can be phased with the Readbackphasing tool, and we used Beagle to phase the remaining SNPs based on population frequencies of haplotypes. Beagle uses a localized haplotype-cluster model, this models haplotype frequencies on a local scale. This takes into account that correlation between markers is a local phenomenon and that LD decays with distance (Browning and Browning 2007; Browning and Browning 2007).

Data filtering

Two datasets with different filtering options were created from the final VCF with VCFtools (Danecek et al. 2011). The first dataset contains all *Hogna* species, which will be further on referred to as the 'complete dataset '. For the second dataset, the species *H. albemarlensis* and *H. jacquesbreli* were excluded as these species are only distantly related to the species group that underwent a parallel radiation at Galápagos. As suggested by De Busschere et al. (2010), the presence of these species at Galápagos may be the result of separate colonization events. Because of their old shared ancestry with the other species, both species groups have only few tags in common, which strongly reduces the number of tags that are shared among all the species. The dataset that excludes these two species is further referred to as the 'galapagoensis clade' dataset, and includes the species *H. espanola, H. snodgrassi, H. junco, H. hendrickxi* and *H. galapagoensis*. The following filtering options were used for both datasets: a minor allele frequency of 0.0125, maximum allele frequency of 0.9875, maximum 2 alleles per SNP and a minimum read depth of 10. The filtering of a minimum genotype quality of 20 was already done during phasing. For the complete dataset and 'galapagoensis clade' dataset, we further specified that 95% and 80% of the SNPs must be present in the individuals respectively.

4.3. Population structure

PCoA

Population structure was first visually inspected by conducting a Principal Coordinates Analysis (PCoA) based on the pairwise, individual-by-individual genetic distance for codominant data with Genalex v6.502 (Peakall and Smouse 2012). The genetic distance matrix is calculated for each locus, as a set of squared distances defined as d2(ii, ii) = 0, d2(ij, ij) = 0, d2(ii, ij) = 1, d2(ij, ik) = 1, d2(ij, kl) =2, d2(ii, jk) = 3, and d2(ii, jj) = 4 , with ith, j-th, k-th and l-th different alleles (Smouse and Peakall 1999). Al the distances matrices of the loci are summed for, to get the overall individual-by-individual genetic distance matrix. Considering

that SNPs located within the same RADtag are linked to each other, we selected one random SNP per tag using a custom Python script.

Cluster analysis

We determined the patterns of genetic structure between the different populations by inferring the number of genetic clusters that best fits the data. A Bayesian analysis is used, which assigns individuals to clusters (K) by detecting allele frequency differences, which was implemented in STRUCTURE (Porras-Hurtado et al. 2013; Pritchard, Stephens, and Donnelly 2000). Also for this analysis, one random SNP per RADtag was selected. For the complete dataset, we ran ten independent runs for each K-values, and K-values ranging from 2-8 with 100.000 Markov chain runs and with a burnin of 20.000. For the 'galapagoensis clade' dataset, 9-11 independent runs were executed with K ranging from 2-6 with 300.000 Markov chain runs and with a burnin of 60.000. Outputs are processed using harvester (Earl and vonHoldt 2012). For each K-value, the clustering of the populations was assessed and runs with the same clustering scenario are processed to a mean matrix with the program CLUMPP v1.1.2. (Jakobsson and Rosenberg 2007) using the Greedy algorithm. The results were visualised using Distruct v1.1 (Rosenberg 2004).

Population differentiation

To assess the genetic differentiation between the populations, the fixation index (F_{st}) between all species was calculated using Genepop v4.4 (Rousset 2008) for both datasets.

4.4. Genomic divergence

The F_{st} (Weir) of each individual SNP for intra and inter island comparisons are calculated. To determine if the same SNPs are strongly differentiated within and between islands. The following population comparisons were analysed, within islands: H. junco - H. snodgrassi (San Cristóbal), H. hendrickxi - H. galapagoensis (Santa Cruz), and between islands: H. hendrickxi - H . snodgrassi ('low' ecotype) and H. galapagoensis - H. junco ('top' ecotype) were calculated using VCFtools.

To identify candidate loci that might potentially be associated with sites subjected to natural selection, we need to assess which loci are significantly differentiated, i.e. outlier loci. Two models are defined to identify these loci, the null model without selection and the alternative model with selection. Selection is modelled by decomposing the locus-specific F_{st} coefficient into a population-specific component, shared by all loci, and into a locus specific component, shared by all populations. The model without selection only includes the population-specific component and the model with selection includes both population-specific component and a locus-specific component. The Posterior Odds (PO) are determined, which is the ratio of the posterior probabilities of the alternative model compared to the null model. This allows the control of the False Discovery Rate (FDR), which represents the expected proportion of false positives among outlier markers. BayeScan v2.1 (Foll and Gaggiotti 2008) uses this Bayesian approach. The outlier loci were assessed for the following pairs: San Cristóbal (H. *junco* - H. *snodgrassi*), Santa Cruz (H. *hendrickxi* - H. *galapagoensis*), 'low' ecotype (H. *hendrickxi* - H . *snodgrassi*), 'top' ecotype (H. *galapagoensis* - H. *junco*). The burnin was set to 50.000 with a thinning interval of 10. The sample size was 5000. There were 20 pilot runs, each with a length of 5000. The prior odds had the default value of 10, which means that that prior likelihood of the neutral model is 10 times higher than the alternative model with selection. The FDR of 0.05 was used to assess the outlier loci.

4.5. Phylogenetic relationship

We used two approaches to gain insight into the phylogenetic relationships among the different species. First, we integrated the phylogenies obtained from separate RADtags into a single consensus phylogeny by means of *BEAST (Bayesian Evolutionary Analysis by Sampling Trees) (Heled and Drummond 2010). *BEAST uses a Bayesian Markov chain Monte Carlo method for estimating the species tree from multilocus data. As input, we selected the 15 longest RADtags from the complete dataset and specified a Hasegawa-Kishino-Yano (HKY) substitution model for all fragments, a strict clock and a 'Yule ' tree prior . The MCMC chain length was set to 97.734.000 generations. After discarding the first 35.000.000 states, Tracer V1.6 (Rambaut and Drummond 2004) was used to inspect the likelihood and the Effective Sample Size (ESS) of the parameter estimates. The consensus tree with the highest posterior probability was selected with TreeAnnotator v1.8.3 (Drummond, Rambaut & Suchard 2002-2016) and the same program was used to calculate the posterior probabilities of the nodes. The final tree was visualized with Figtree v1.4 (Rambaut 2006).

Second, we attempted to visualize the phylogenetic relationships obtained by constructing individual trees per RADtag with RAxML (Randomized Axelerated Maximum Likelihood; Stamatakis 2014). Two sets were used in this analysis. One set, wherein all species were included, contained 266 RADtags and was used to infer general relationships among all the species. A second set only contained the species of the 'galapagoensis clade' and contained only those RADtags containing an outlier SNP in the highland/lowland comparison. Phylogenies of this set were used to gain insight into the evolutionary history of ecotypic differentiation within this parallel radiation. More specifically, we here test if for these outlier loci a clustering is observed with respect to their ecotype.

Both datasets were created from the original dataset to retain as much information as possible. For the first dataset, four individuals per species (except two for individuals H. *albemarlensis*) were selected with the lowest proportion of missing tags. Filtering settings used were minimum allele frequency of 0.0125, maximum allele frequency of 0.9875, proportion of missing data of 95%, mean depth value of 10. Only RADtags with a minimum of four SNPs per tag were selected.

The second dataset, included 5 individuals per species with the least amount of missing data. The filtering settings used were: minimum allele frequency of 0.068, maximum allele frequency of 0.932, proportion of missing data of 95%, mean depth value of 10. Only RADtags with a minimum of 4 SNPs were selected. RADtags for which data was completely absent for one or more individuals were discarded.

The substitution model used was GTR with gamma model of rate heterogeneity. Out of a total of 100 trees, the tree with the highest maximum likelihood was chosen.

5. Results

5.1. Sequencing

Reads

The HiSeq run generated 205.907.374 reads of which 33,45% were discarded due to low quality (9,9%), ambiguous barcodes (11,7%), ambiguous RAD-tags (2,25%) or lack of a complementary read in the forward or reverse read (9,6%) (Figure 4).

Number of reads per individual ranged from 671.992 to 4.848.728, except for four individuals, for which only 41.638 to 65.584 reads were sequenced (Figure 5). After eliminating these latter four individuals we retained 136.704.166 reads in total. Hereafter, we removed 46,4% of the reads, as they constituted PCR duplicates, resulting in a final set of 63.419.370 reads.

The MiSeq run produced 10.277.158 reads, 20,4% were discarded due to an ambiguous barcode, low quality (1,7%), ambiguous RAD-tags(1,3%) or lack of a complementary read in the forward or reverse read (2,7%). In total, 7.577.512 reads were retained, while reads per individual ranged between 244.034 and 792.780 (Figure 5).

Genome assembly & mapping

The assembly of the reference tags resulted in 3.526.761 bp and 63.076 RADtags with an N50 of 284bp (Figure 6, Figure 7) and mean length of 388bp.

Genotype calling & phasing

The Readbackphasing and Beagle algorithm removed respectively 11.446 bp (low genotype quality) and 44.625 bp (missing data, i.e. positions for which no data is available).The phasing resulted in 1.2% of the genotypes phased by Readbackphasing tool, 73.54% by Beagle and the remaining 25,26% of the data is missing. After completion of this genotype calling pipeline, 3.738.330 bp spread over 9.100 RADtags were retained.

Filtering datasets

After filtering, the 'complete' dataset contained 9.172 SNPs located on 500 RADtags with an average 16,8 (SD=21,76) SNPs per RADtag. The 'galapagoensis clade' dataset comprised 46.605 SNPs on 2.773 RADtags with an average of 18,3 (SD=20,12) SNPs per RAD tag.

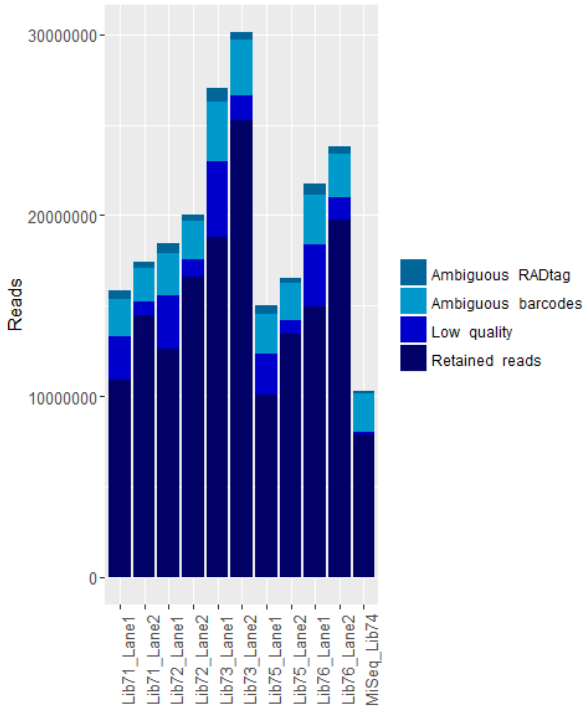


Figure 4-Total obtained reads per lane and library for the MiSeq and HiSeq illumina runs.

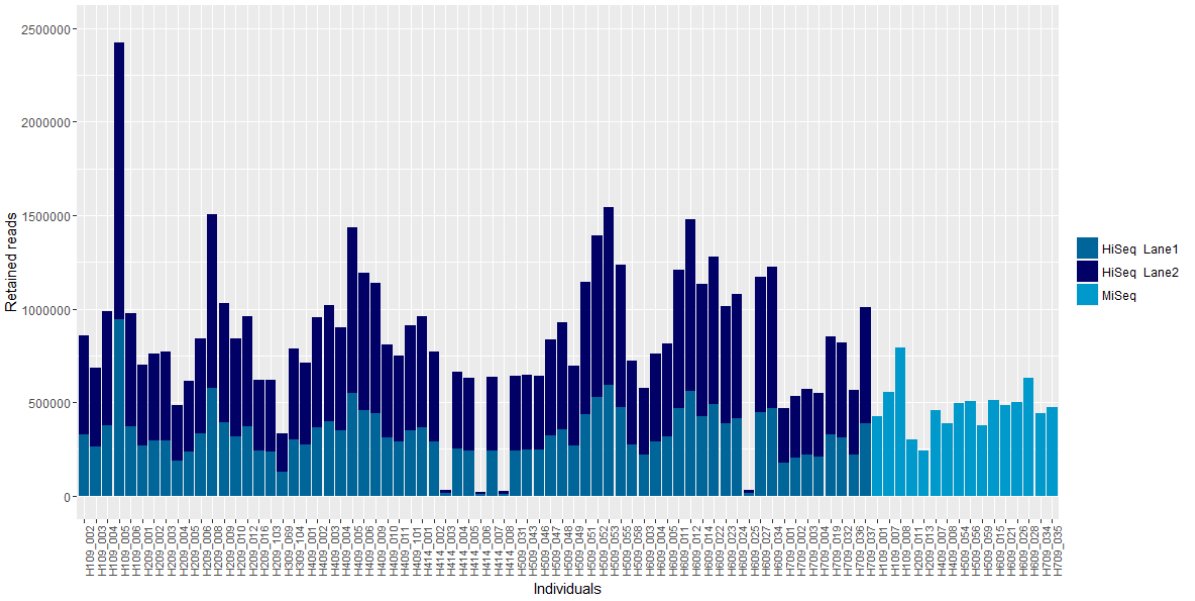


Figure 5-Reads per individual obtained by HiSeq and MiSeq illumina platform.

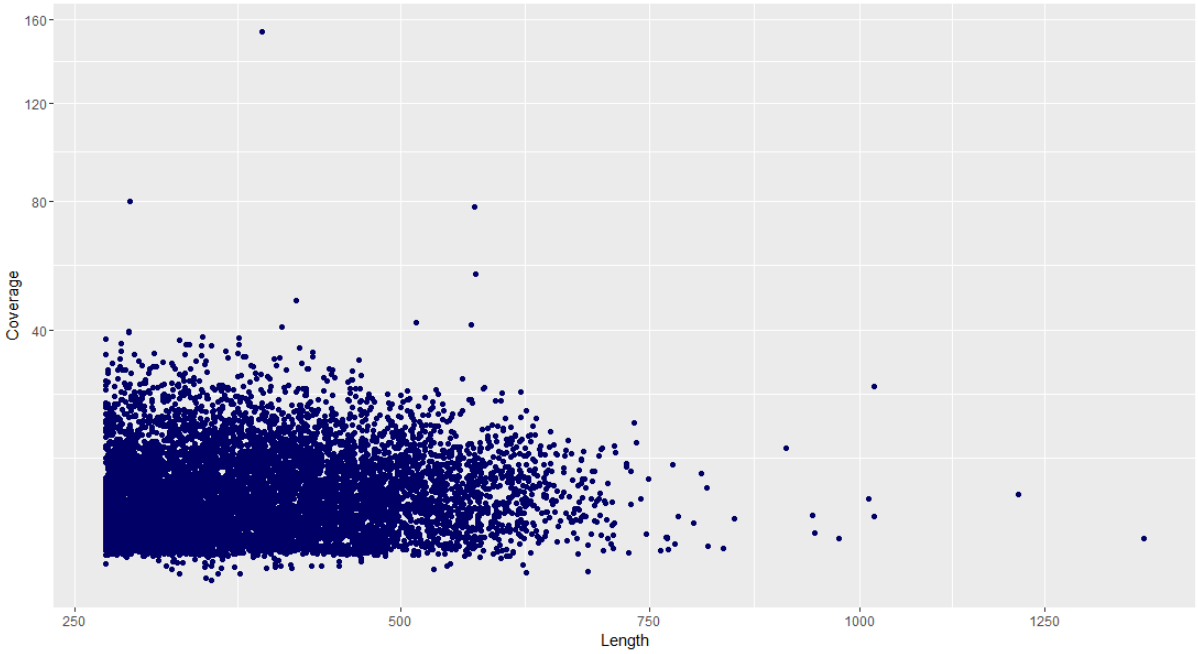


Figure 6–Distribution of coverage and length of reference RADtags of assembly of H. *galapagoensis* individual (H409_008).

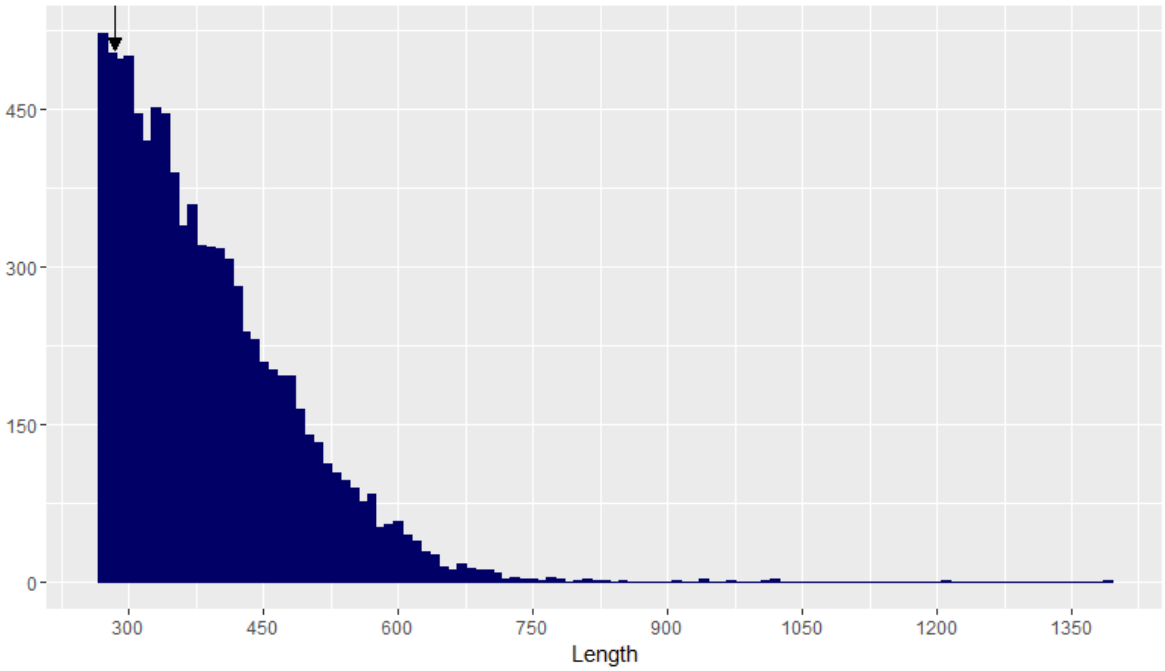


Figure 7-Distribution of the length of the assembled reference RADtags of H. *galapagoensis* individual (H409_008) . Arrow points to N50 of 284

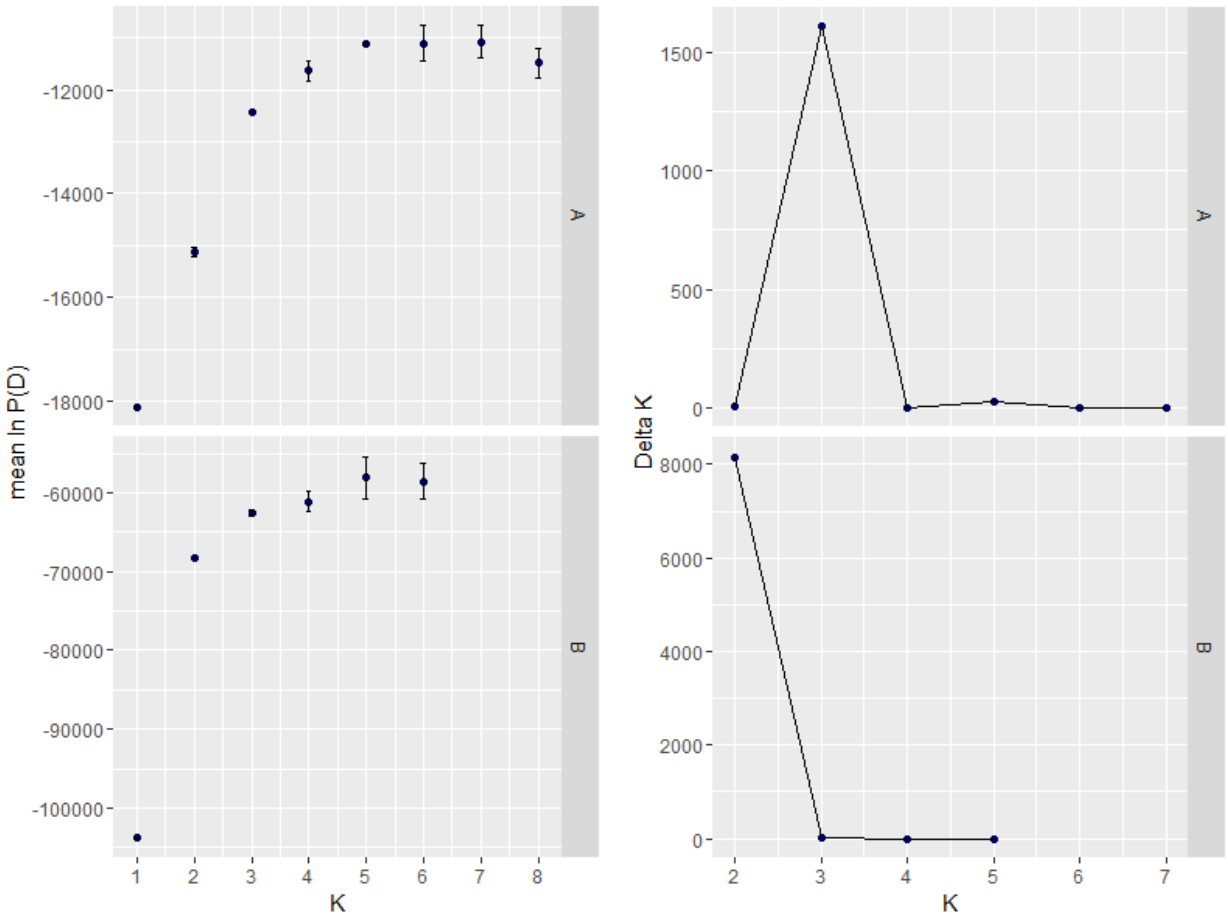


Figure 8-Mean posterior probabilities with SD flags and delta K value of evanno for the complete dataset (A) and the 'galapagoensis clade' dataset (B).

5.2. Population structure

In total, 2.773 SNPs for the 'galapagoensis clade' dataset and 500 SNPs for the 'complete' dataset were considered for the STRUCTURE and PCoA analysis.

Cluster analysis

For the complete dataset and the 'galapagoensis clade' dataset do respectively 3 or 2 clusters best fit the data (Figure 8). This according the Delta K value of evanno we can see that this is the beginning of an asymptote for the ln P(D) value and has a very low standard error.

Clustering of genotypes at various K values was not unequivocal, as on numerous occasions distinct competing scenarios appeared equally likely (Figure 9). For K = 2 we could identify 3 distinct clustering scenarios, differing in the extent to which other populations grouped together with H. *jacquesbreli* and H. *albemarlensis* ,i.e. a) in the first scenario H. *jacquesbreli* and H. *albemarlensis* constituted a single distinct cluster, while the entire galapagoensis clade grouped into a second one (5 runs), b) in the second scenario H. *jacquesbreli* and H. *albemarlensis* grouped together with *H. galapagoensis* (Isabela & Santa Cruz) and H. *hendrickxi* (2 runs) and c) in the third scenario H. *jacquesbreli,* H. *albemarlensis* clustered together with H. j*unco*, H. *snodgrassi* and H. *espanola* (3 runs). For K = 3, all runs showed a consistent outcome in which genotypes of the galapagoensis clade clustered according to their geographical distribution, i.e. a H. *albemarlensis* and H. *jacquesbreli* cluster, an Isabela-Santa Cruz cluster (H. *galapagoensis,* H. *hendrickxi*) and a San Cristóbal-Española cluster (H. *snodgrassi*, H. *junco*, H. *espanola*). Allowing the algorithm to form one extra cluster (K=4) resulted in a) a distinction between H. *jacquesbreli and* H. *albemarlensis* (9 runs) or b) a separation between the San Cristóbal and Española populations (1 run). For K of 5, 6, 7 and 8, the majority of runs corroborated previous results, demonstrating four distinct clusters, namely a H. *jacquesbreli, a* H. *albemarlensis*, an Isabela-Santa Cruz and a San Cristóbal-Española cluster. Alternative scenarios consisted of an additional separate clustering of the San Cristóbal and Española populations.

The clustering analyses of the 'galapagoensis clade' dataset (Figure 10) was once more not unequivocal at various different K values, as several scenarios are equally as likely. For K = 2, the clustering was consistent for all the runs in which clustering occurred according geography corroborating previous results, i.e. an Isabela-Santa Cruz cluster (H. *galapagoensis,* H. *hendrickxi*) and a San Cristóbal-Española cluster (H. *snodgrassi*, H. *junco*, H. *espanola*). For K values 3, 4, 5 and 6, the Isabela-Santa Cruz cluster is consistently present, while various clustering occurs within the Isabela-Santa Cruz cluster.

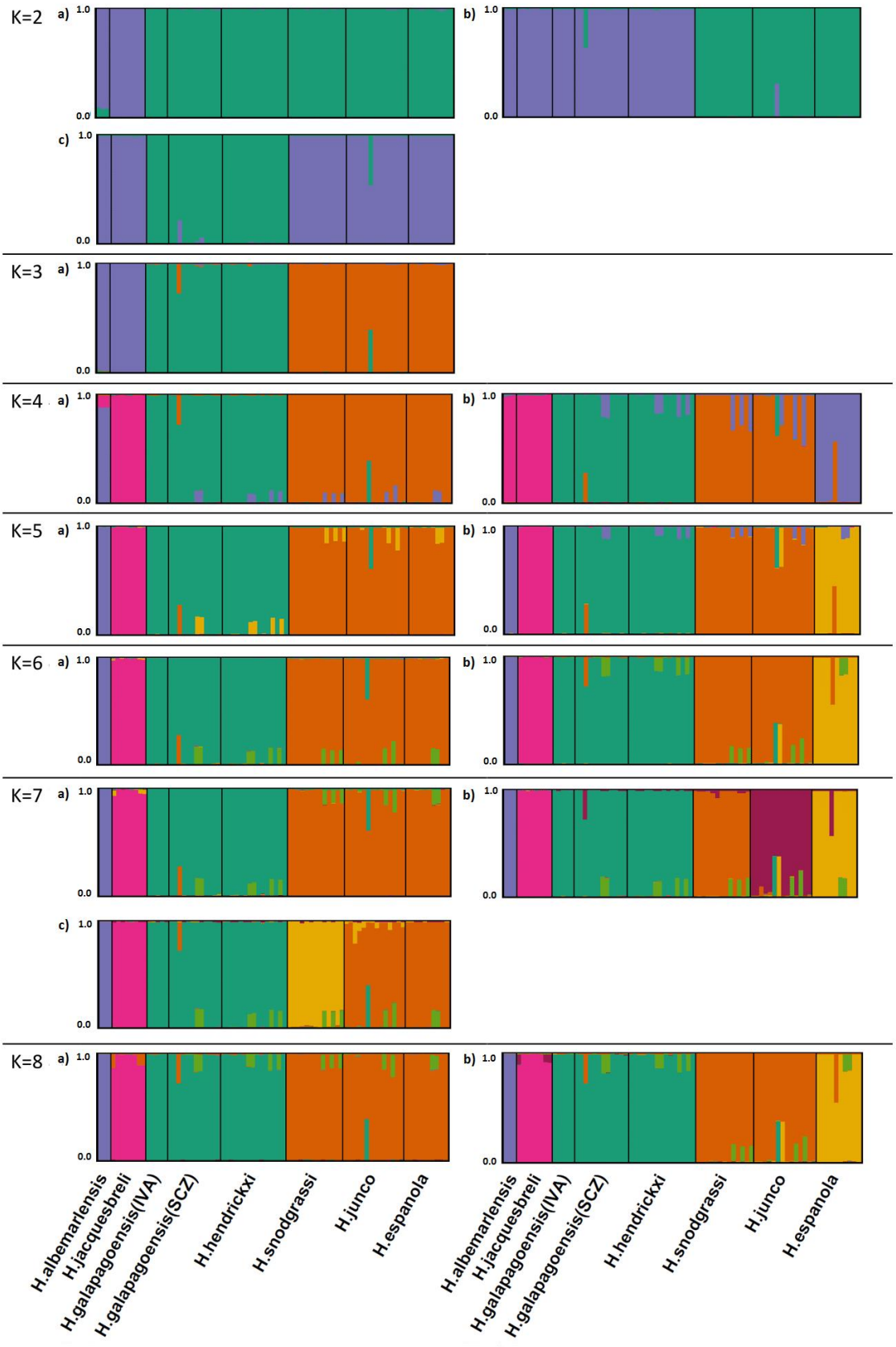


Figure 9- Population structure for SNPs of dataset with all species. Scenarios were present in the following runs: K2: (a) 5 runs, (b) 2 runs and (c) 3 runs. K3: (a) 10 runs. K4: (a) 9 runs and (b) 1 run. K5: (a) 9 runs and (b) 1 run. K6: (a) 8 runs and (b) 2 runs. K7: (a) 8 runs, (b) 1 run and (c) 1 run. K8: (a) 9 runs and (b) 1 run.

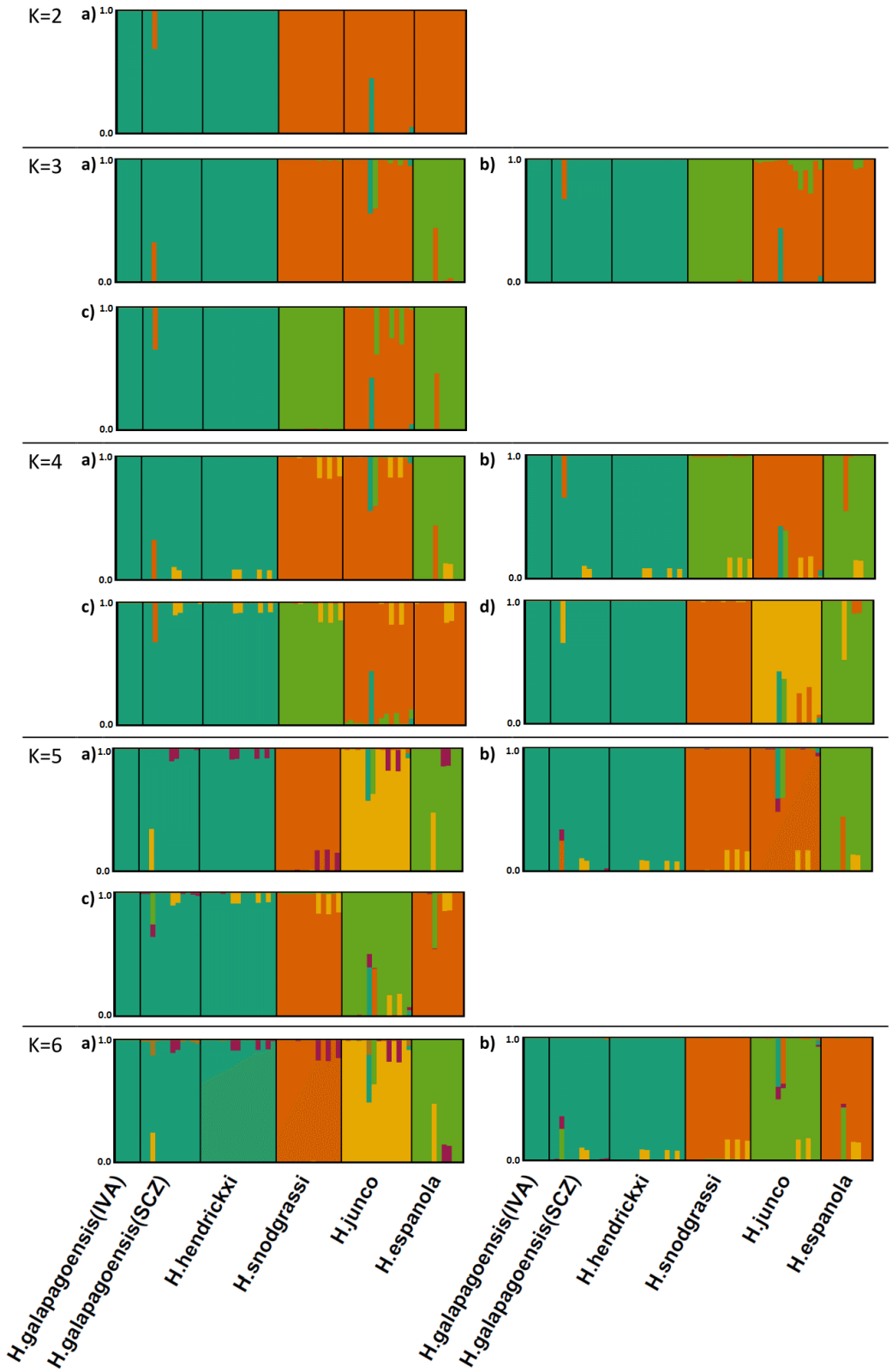


Figure 10-Population structure for SNPs of 'galapagoensis clade' dataset. Scenarios were present in the following runs: K2: (a) 10 runs. K3: (a) 8 runs, (b) 3 runs and (c) 1 run. K4: (a) 5 runs, (b) 2 runs, (c) 2 runs, (d) 1 run. K5: (a) 7 runs, (b) 1 run and (c) 1 run. K6: (a) 10 runs and (b) 1 runs.

PCoA

In the PCoA of the complete dataset (Figure 11), corroborating previous results of the STRUCTURE analysis, we can identify four different groups (equal to K=4, scenario a). Coinciding with scenario a of K value 4, we observe an H. *albemarlensis* cluster, an H. *jacquesbreli* cluster, the Isabella-Santa Cruz cluster and the Española- San Cristóbal cluster.

The PCoA of the 'galapagoensis clade' dataset (Figure 112) shows similar results as the STRUCTURE analysis, we can differentiate two clusters according the first axis. The first cluster consists of H. *galapagoensis* and H. *hendrickxi* and the second cluster consists of H. *snodgrassi*, H. *junco* and H. *espanola*. No differentiation can be made according the second axis between the different clusters. Within these groups, there is no further clustering of the species. The third axis displays a differentiation of H. *espanola* from H. *junco* and H. *snodgrassi*.

There are two individuals from H. *junco* and H. *galapagoensis* (Santa Cruz) that lay in between the two clusters in both PCoA's.

Population differentiation

An F_{ST}-heatmap showed strong genetic differentiation between most of the population pairs (Figure 13). H. *albemarlensis* and H. *jacquesbreli* are the strongest differentiated from the 'galapagoensis clade' species, in which H. *albemarlensis* has the highest pairwise F_{st} with the other species. They are also strongly differentiated from each other with an F_{st} of 0.3956. Similar to the Bayesian clustering analysis and PCoA, we identified two clusters within the 'galapagoensis clade'. The F_{st}-values within the Isabela-Santa Cruz cluster has a range of 0.0148-0.0479 and the Española- San Cristóbal cluster has a range of 0.0572-0.0929. The between F_{st} -values between both clusters range from 0.1535-0.2084.

For the 'galapagoensis clade' dataset (Figure 14), we yet again see similar results to the Bayesian clustering analysis and Principal Coordinate analysis, in which two clusters emerge according to their geography. The F_{st} -values within the Isabela-Santa Cruz cluster are ranging between 0.036-0.1029 and the Española- San Cristóbal cluster have a range of 0.109-0.1736. The F_{st} -values between both clusters range from 0.3063 to 0.3923.

The F_{st}-values of the 'galapagoensis clade' dataset are approximately twice as high when compared with the pairwise F_{st}-values of the complete dataset. But for both datasets, the strongest differentiation can be observed within the Española- San Cristóbal cluster.

5.3. Genomic divergence

The F_{st} between the species of the same islands are 0.109 for San Cristóbal (H. *junco* - H. *snodgrassi*) and 0.036 for Santa Cruz (H. *hendrickxi* - H. *galapagoensis*). The F_{st} between the same ecotype of the different islands is 0.3447 for the 'low' ecotype (H. *hendrickxi* - H. *snodgrassi*) and 0.3063 for the 'top' ecotype (H. *galapagoensis* - H. *junco*). When we look at the distribution of F_{st} -values across SNPs (Figure 15), we can see that overall the SNPs have a very low F_{st}-value. The histograms of San Cristóbal and Santa Cruz have a thin tail, and accordingly a low amount of SNPs with a high F_{st}. The histograms of the F_{st} -values of the top ecotype and low ecotype comparison have a wider tail with a small increase towards a high F_{st}. On the scatterplot of the F_{st}-values (Figure 16), we can observe that the San Cristóbal comparison has more SNPs with a high F_{st}-value in contrast to the Santa Cruz comparison. The top ecotype and low ecotype comparisons display

higher F_{st}-values, which shows divergence is higher between the same ecotypes in contrast to divergence within islands.

The comparison of the F_{st} of top ecotype and the 'low' ecotype (Figure 17) shows that the same SNPs have a similar F_{st} distribution, showing the majority of SNPs having a low F_{st}-value, while a smaller portion has a higher F_{st}-value. Many of these SNPs are shared between islands. In contrast, comparing the F_{st} distribution of San Cristóbal and Santa Cruz (Figure 18) shows that several SNPs are characterized by strong within-island genetic differentiation, yet none of these outlying SNPs are shared across islands.

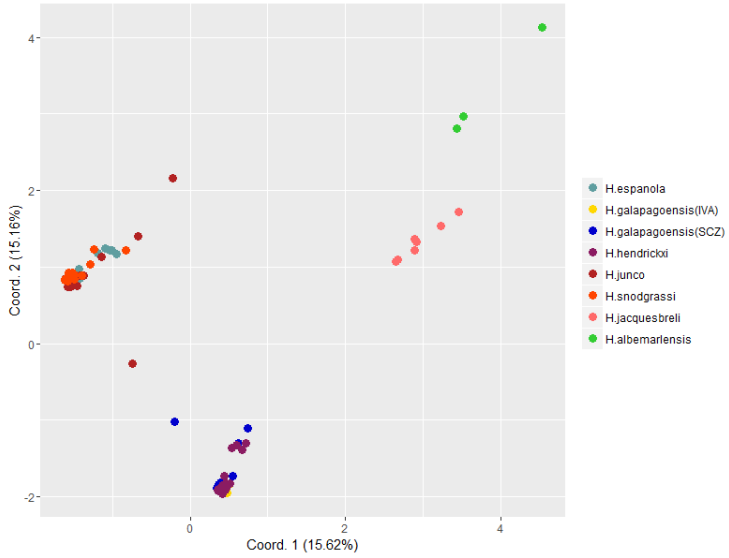


Figure 11-Principal Coordinate Analysis (PCoA). PCoA for all species including 500 polymorphic markers.

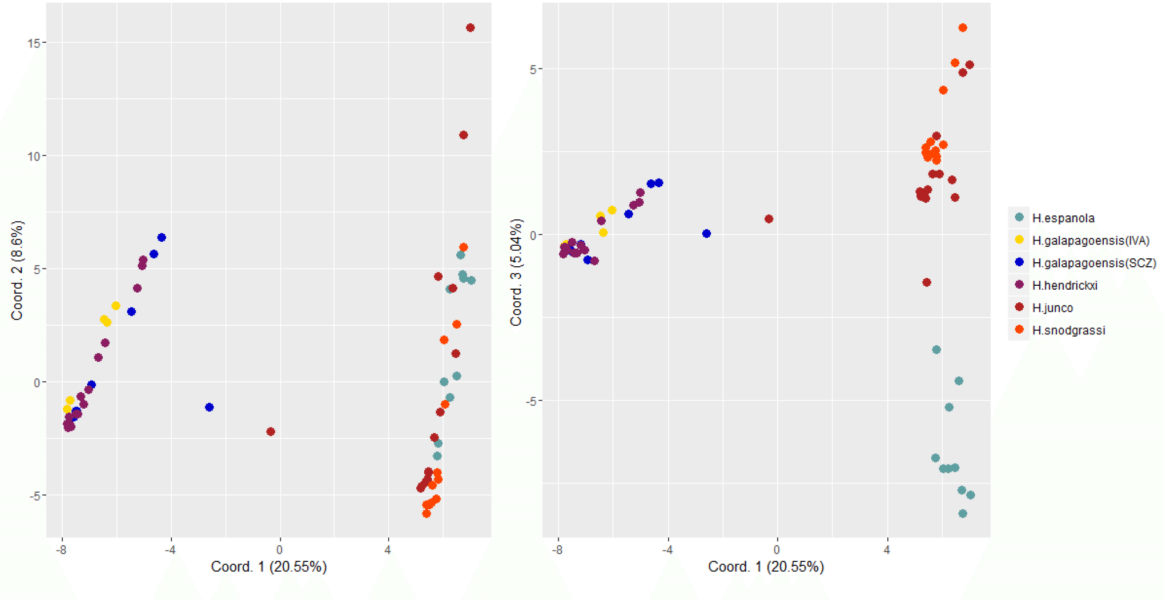


Figure 12-Principal Coordinate Analysis (PCoA). PCoA for 'galapagoensis clade' species including 2.773 polymorphic markers.

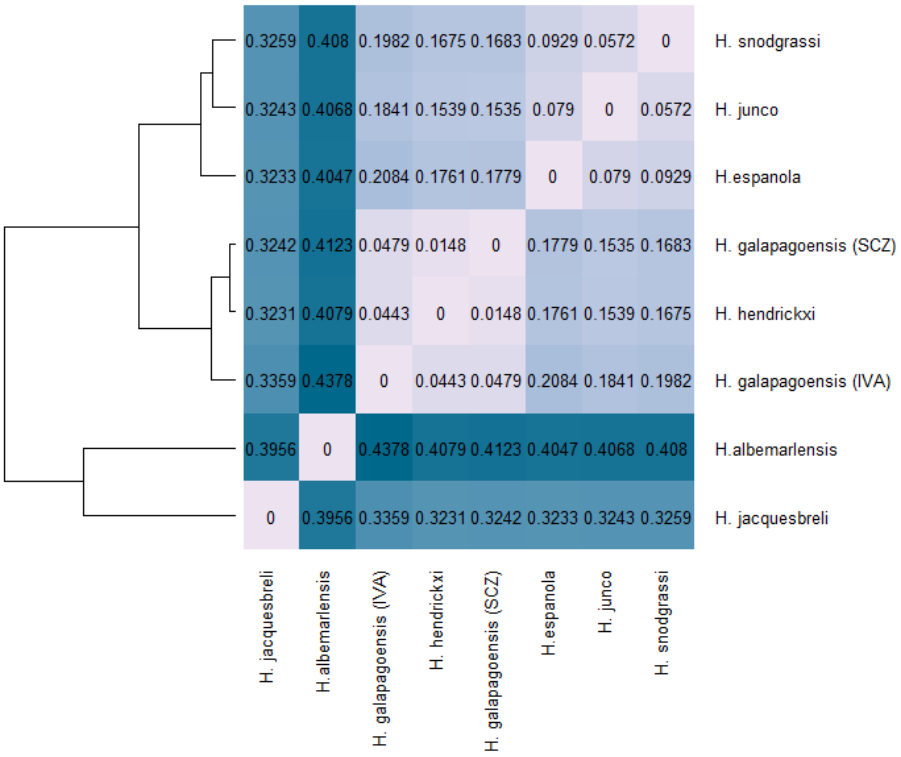


Figure 13-Pairwise F_{st} -values of all the species.

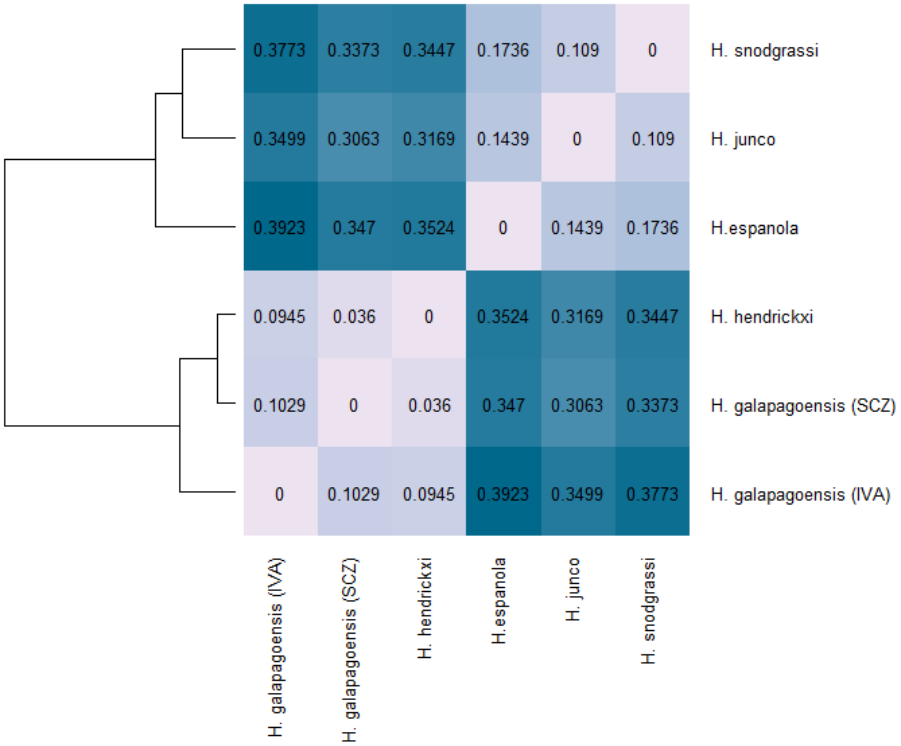


Figure 14- Pairwise F_{st} -values of the species belonging to the 'galapagoensis clade'.

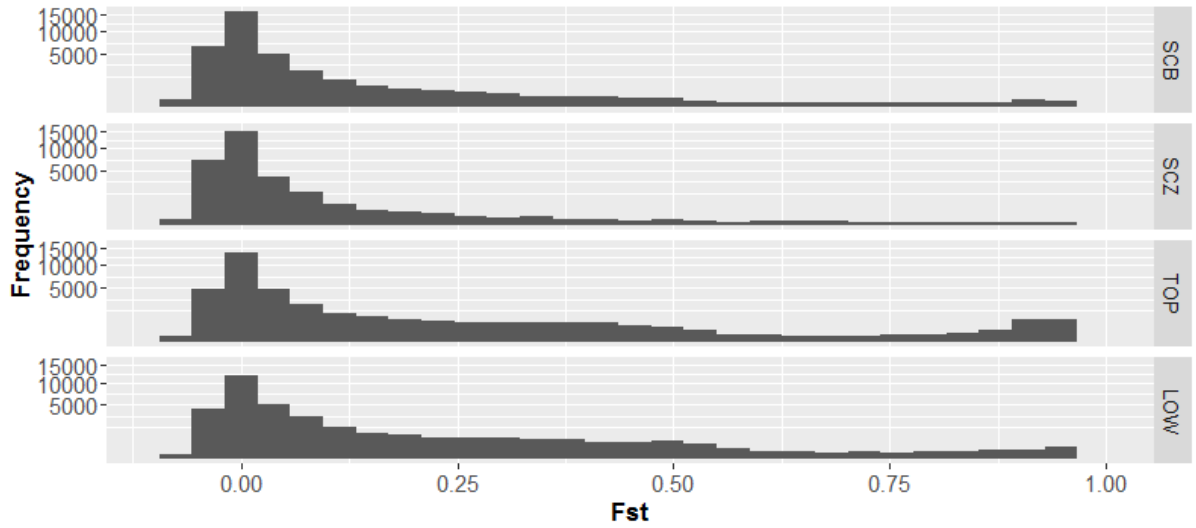


Figure 15. Histogram of F_{st}-values with SCB (San Cristobal: H. *junco* – H. *snodgrassi*), SCZ (Santa Cruz : H. *hendrickxi*-H. *galapagoensis*), LOW (H. *hendrickxi*-H . *snodgrassi*), TOP (H. *galapagoensis*-H. *junco*).

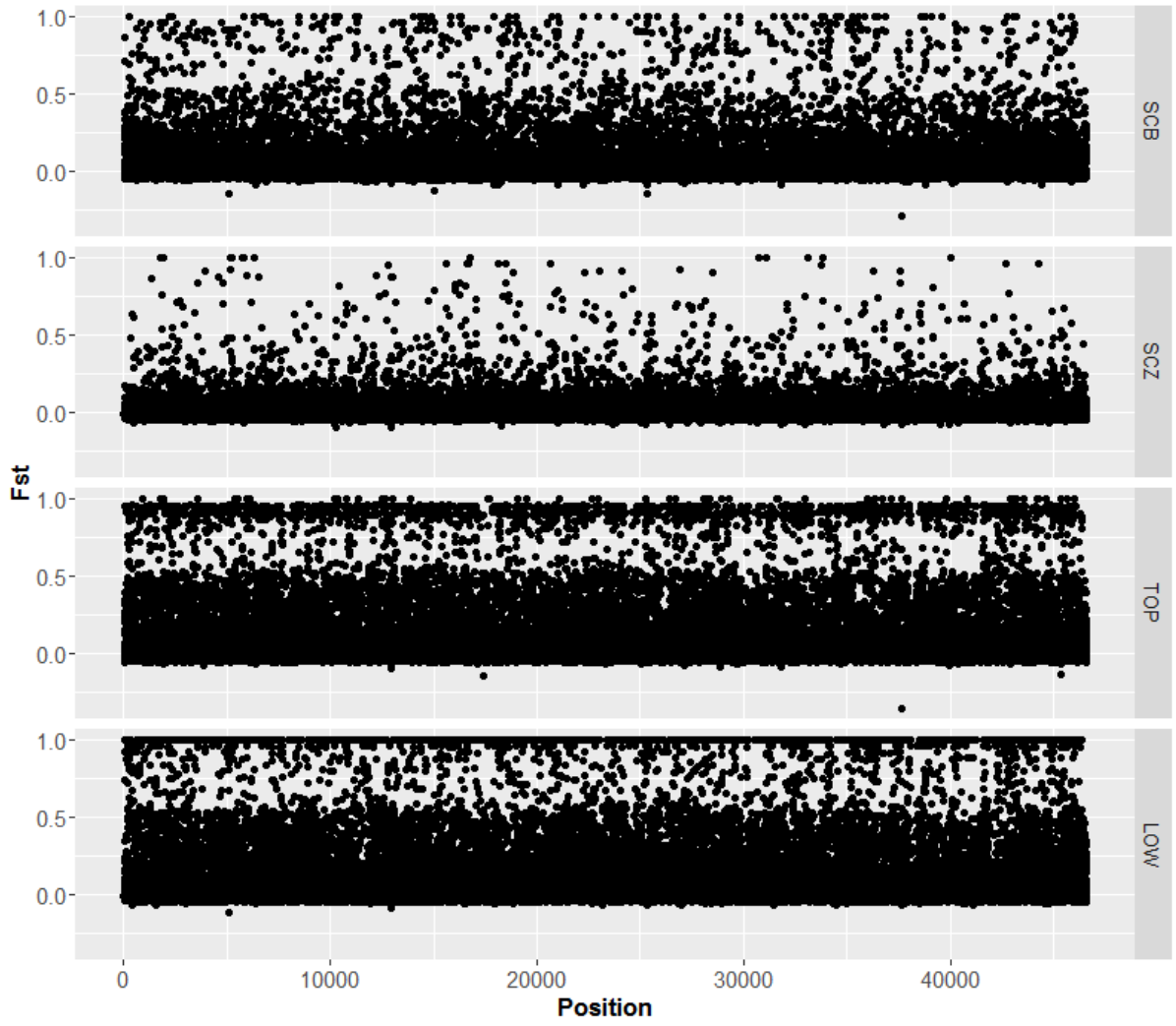


Figure 16. Distribution of F_{st}-values of SCB (San Cristobal: H. *junco* – H. *snodgrassi*), SCZ (Santa Cruz : H. *hendrickxi*-H. *galapagoensis*), LOW (H. *hendrickxi*-H . *snodgrassi*), TOP ('top' ecotypes: H. *galapagoensis*-H. *junco*).

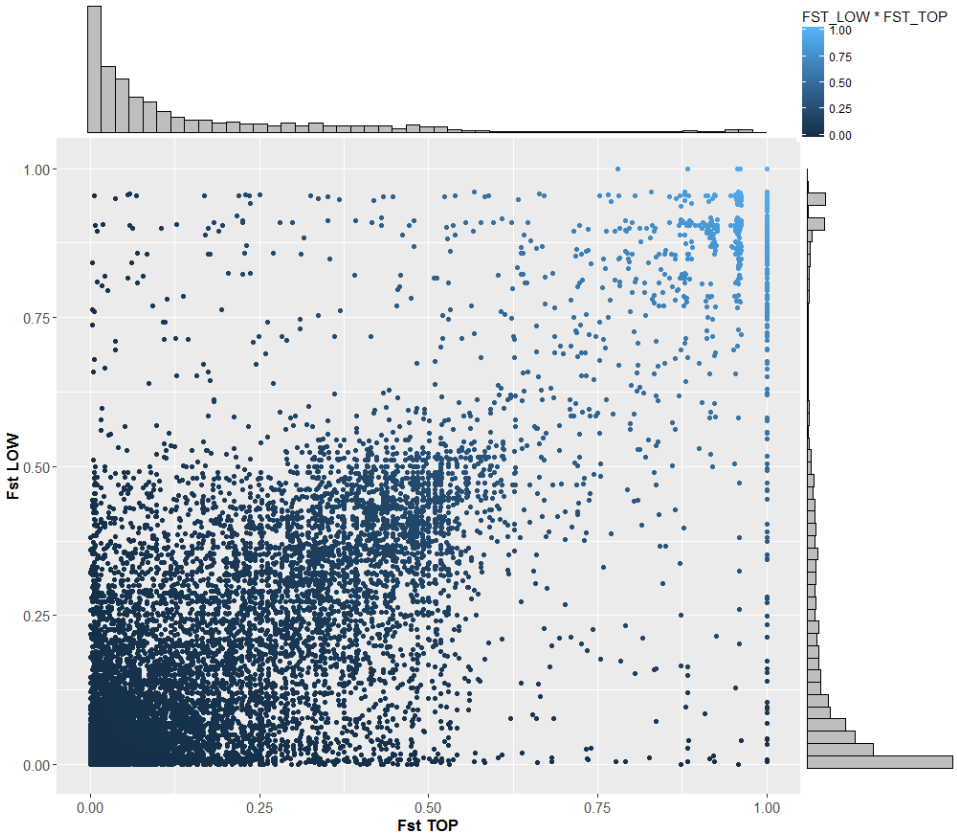


Figure 17-Plot of the F_{st}-values between low ecotype (LOW: H. *hendrickxi*-H . *snodgrassi*) against the F_{st}-values between the 'top' ecotypes (TOP: H. *galapagoensis*-H. *junco*) of Santa Cruz and San Cristóbal.

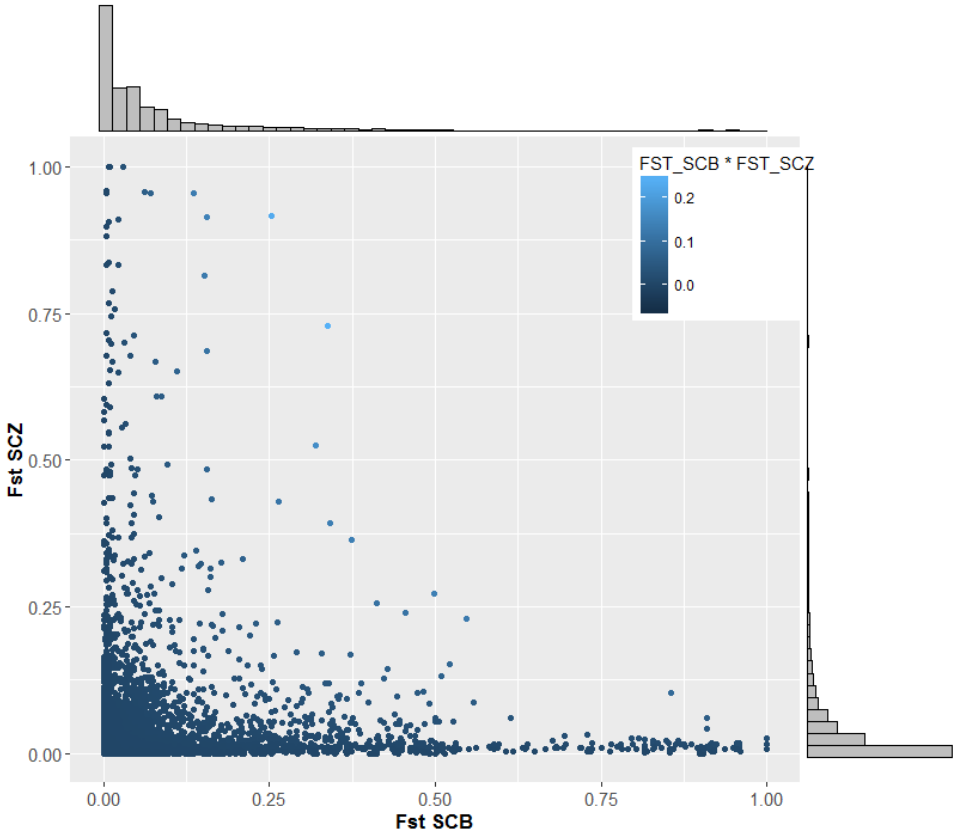


Figure 18- Plot of the F_{st}-values between San Cristóbal species species(H. *junco* – H. *snodgrassi*) against the F_{st}-values between the Santa Cruz (H. *hendrickxi*-H. *galapagoensis*).

Outlier analysis

Prior to each analysis, monomorphic sites were discarded, resulting in the following number of SNPs for the within island comparison: 35.558 for H. *junco* – H. *snodgrassi* (San Cristóbal), 30.382 for H. *hendrickxi*-H. *galapagoensis* (Santa Cruz), and for the between island comparison: 36.781 H. *hendrickxi*-H . *snodgrassi* (low ecotype) and 36.215 H. *galapagoensis*-H. *junco* ('top' ecotype).

Using a False Discovery Rate of 0.05, 97 (0.27%) outlier SNPs and 1753 (4.77%) ones were identified for respectively H. *galapagoensis* - H. *junco* and H. *hendrickxi* - H .*snodgrassi*. When contrasting different ecotypes within an island, 240 SNPs (0.79 %) were considered as outliers for H. *hendrickxi* - H. *galapagoensis*, while for H. *junco* – H. *snodgrassi* 518 (1.46 %) outliers were identified (Table 1, Figure 19&Figure 20). Corresponding with previous results, the F_{st}-values of Santa Cruz and the 'top' ecotype comparison are higher in contrast to San Cristóbal comparison and the 'top' ecotype comparison.

These highly differentiated SNPs were dispersed across a large number of distinct RADtags in each analysis, namely 195 (Santa Cruz), 413 (San Cristóbal), 48 ('top' ecotype) and 984 ('low' ecotype).

	SCZ	∩	SCB	TOP	∩	LOW
outlier SNPs	240	0	518	97	88	1753
			SCZ∩SCB∩TOP∩LOW			
			0			
RADtags	195	40	413	48	47	984
			SCZ∩SCB∩TOP∩LOW			
			1			

Table 1- Results BayeScan analysis, the outlier loci detected with FDR of 0.05 and the RADtags on which these are located. SCB (San Cristobal: H. *junco* – H. *snodgrassi*), SCZ (Santa Cruz: H. *hendrickxi*-H. *galapagoensis*), LOW (low ecotypes: H. *hendrickxi*-H . *snodgrassi*), TOP (top ecotypes: H. *galapagoensis*-H. *junco*).

5.4. Phylogenetic relationship

Reconstructing the phylogenetic relationship based on the multispecies coalescent revealed multiple highly supported nodes (Fig. 21). However, the effective sample size for the likelihood and posterior probability were lower (52 and 78 respectively) than the recommended 100 (Drummond et al. 2007) and were hardly increased by raising the runtime of the analysis. This suggests that multiple local optima are sampled and results should therefore be interpreted with caution.

At the most basal level, the species H. *jacquesbreli* is separated with high support from the remaining species. At the next level, H. *albemarlensis* is separated from the species in the 'galapagoensis clade', and confirms the treatment of the latter species group, wherein parallel evolution occurred as a monophyletic clade. Within this 'galapagoensis clade', two clusters are observed. One cluster contains the species from the oldest and easternmost islands San Cristóbal and Española. The second highly supported cluster comprises the two

allopatric populations of H. *galapagoensis*, wherein the population from Santa Cruz clusters with weak support with H. *hendrickxi* from Santa Cruz. In sum, this multispecies analysis revealed that highland and lowland species from the same island are monophyletic.

Superimposing the Maximum likelihood trees constructed for the 266 RAD tags of the complete dataset revealed that a very large amount of variation is present among the individual trees (Fig. 22). With the exception of a weak monophyletic signal of H. *jacquesbreli* and the species from San Cristóbal and Santa Cruz, this variation did not allow inferring of consistent phylogenetic patterns among the trees. Superimposing the Maximum likelihood trees constructed for the 36 RADtags containing an outlier SNP (Fig. 23) resulted, in line with the previous analysis, substantial variation among the genealogies, resulting in only weak general phylogenetic patterns. Despite this variation, some clustering of H. *hendrickxi* and H. *galapagoensis* could be distinguished, which suggests that, for these outliers as well, species cluster according to geography rather than to ecotype.

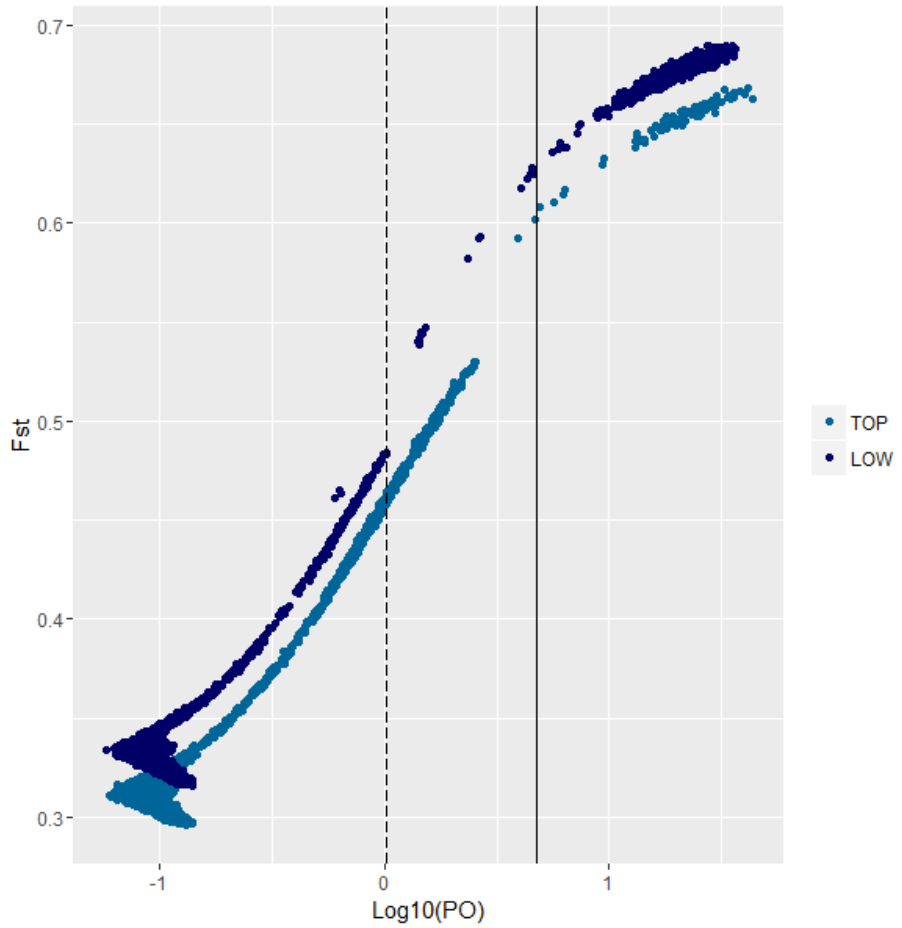


Figure 19. F_{st}-Log10 posterior probability. TOP (top' ecotypes: H. *galapagoensis*-H. *junco*), LOW ('low' ecotype: H. *hendrickxi*-H . *snodgrassi*), FDR value of 0.05 of TOP (full line) and LOW (dashed line).

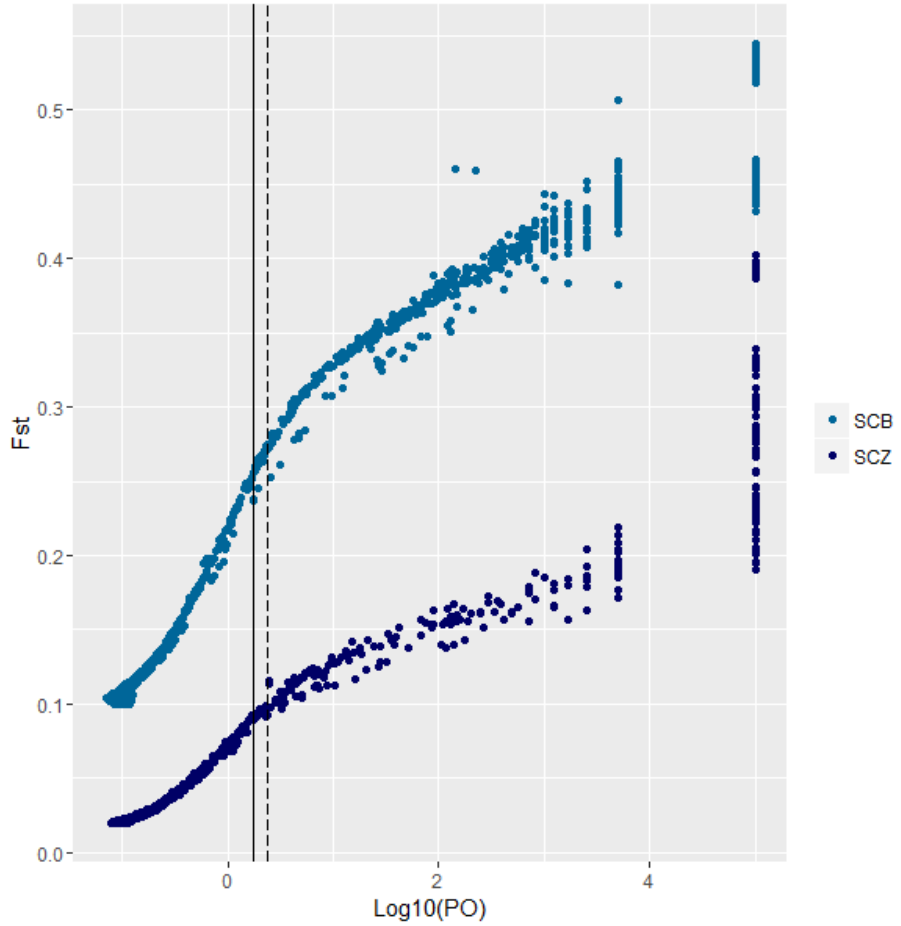


Figure 20- F_{st} -Log10 posterior probability. Value of infinity are set to a value of 5 to make it visual. SCB (San Cristobal : H. *junco* - H. *snodgrassi*), SCZ (Santa Cruz : H. *hendrickxi* - H. *galapagoensis*), FDR value of 0.005 of San Cristobal (full line) and Santa Cruz (dashed line).

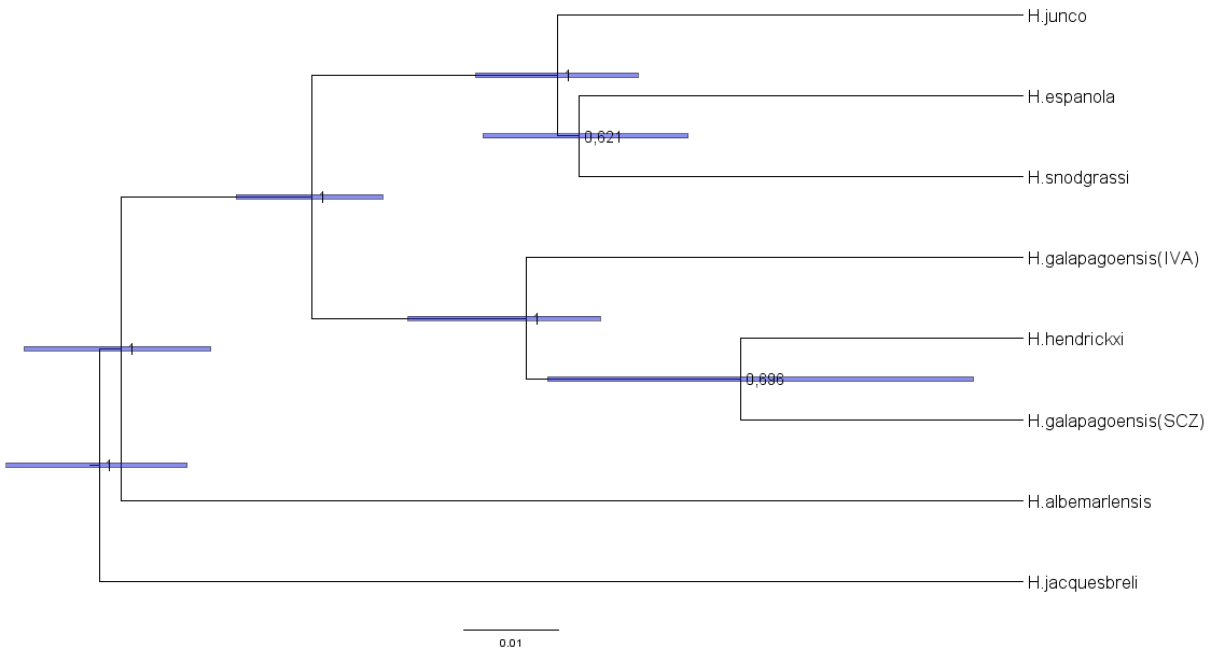


Figure 21-Consensus tree of *beast analysis of 15 RAD tags.

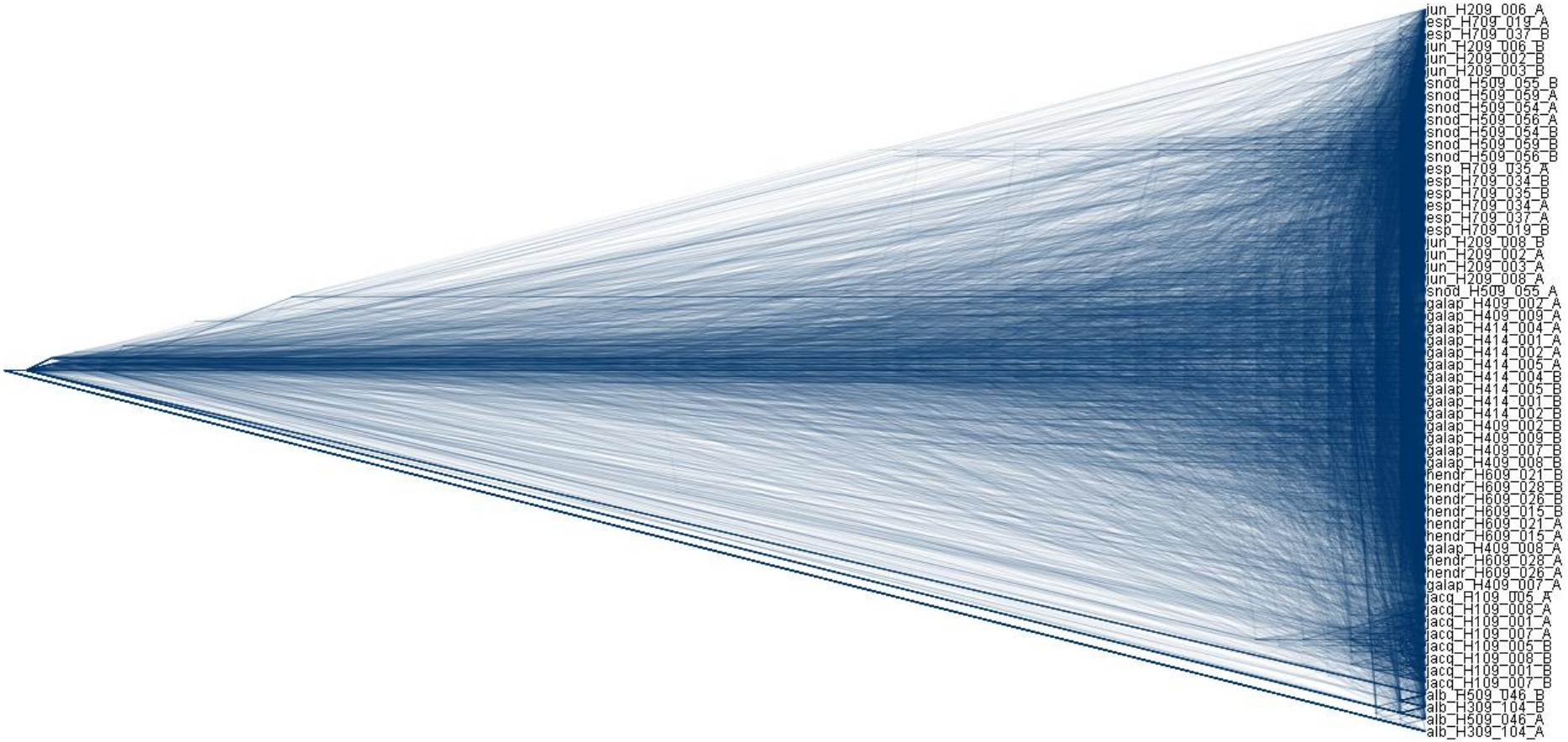


Figure 22-RAxML trees, branch length proportional. Visualisation with Densitree, ordering according closest outside first, trees of 266 neutral RAD tags. H. *junco* (jun); H. *galapagoensis* (galap_H414 occurs on Isabela, galap_H409 occurs on Santa Cruz); H. *jacquesbreli* (jacq), H. *snodgrassi* (snod); H. *hendrickxi* (hendr); H. *espanola* (esp) and H. *albemarlensis* (alb).

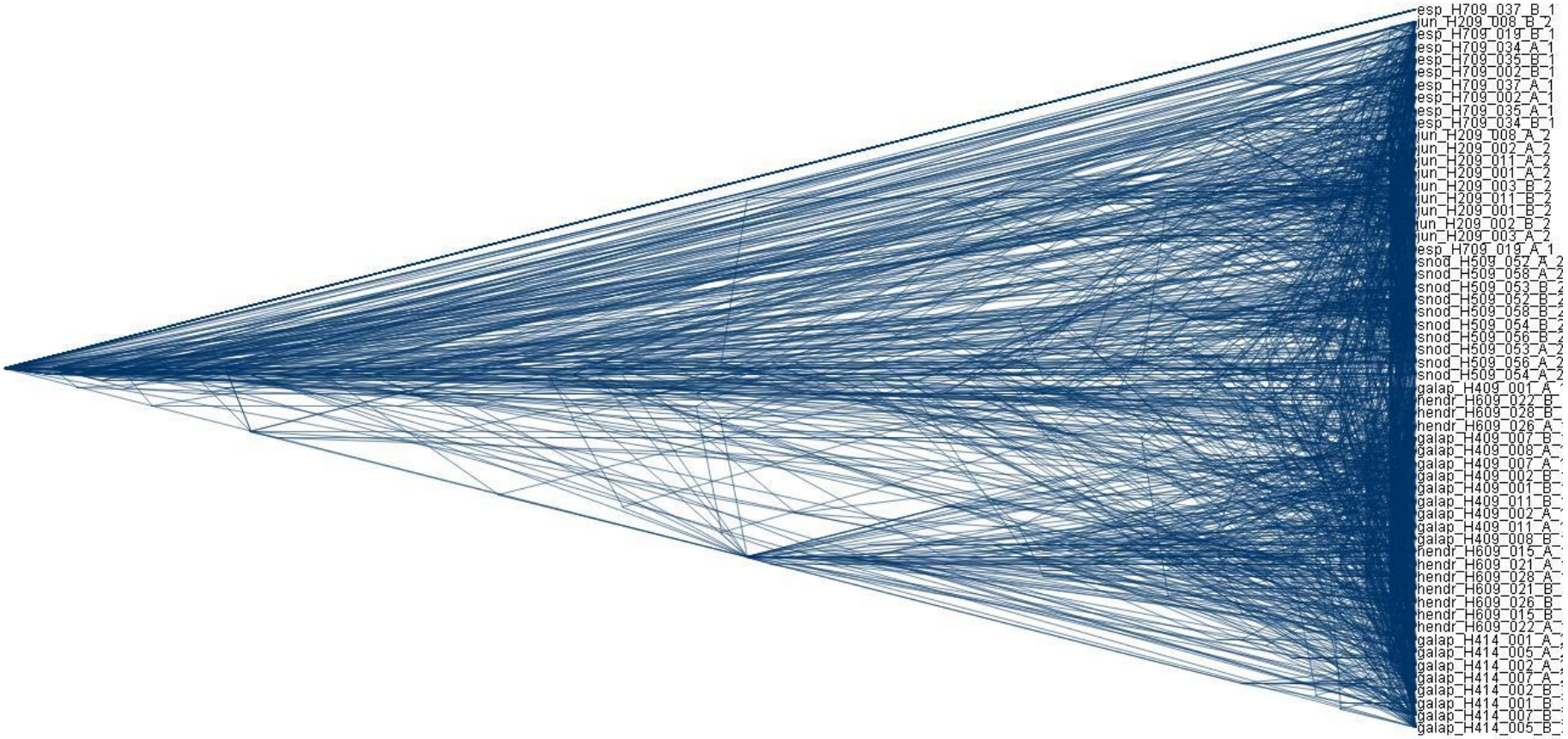


Figure 23- RAxML trees, branch length proportional. Visualisation with Densitree, ordering according closest outside first, trees of 36 RAD tags containing an outlier SNP. H. *junco* (jun); H. *galapagoensis* (galap_H414 occurs on Isabela, galap_H409 occurs on Santa Cruz); H. *snodgrassi* (snod); H. *hendrickxi* (hendr) and H. *espanola* (esp).

6. Discussion

Population structure

The Bayesian cluster analysis, PCoA and pairwise F_{st} estimates gave corresponding results for the population structure among the *Hogna* species. H. *jacquesbreli* and H. *albemarlensis* are strongly diverged from the 'galapagoensis clade' species (H. *galapagoensis*, H. *hendrickxi*, H. *snodgrassi*, H. *junco* and H. *espanola*). The PCoA and the pairwise F_{st} values indicate that H. albemarlensis is the most strongly diverged from the 'galapagoensis clade' species and the divergence between H. *jacquesbreli* and H. *albemarlensis* is approximately equal to the divergence between the galapagoensis clade and H. *jacquesbreli* and H. *albemarlensis*.

The results of our structure analysis differed strongly from the conclusions outlined in De Busschere et al. (2012), where genotypes of COI and 28S rDNA were grouped into seven distinct clusters, representing the seven different species. In contrast, our genome-wide analysis provided most support for a cluster size of three. We can observe that H. galapagoensis and H. hendrickxi are always assigned to the same cluster, while the clustering of H. *snodgrassi*, H. *junco* and H. *espanola* is variable. This indicates that the species of the Isabela-Santa Cruz cluster (H. *galapagoensis,* H. *hendrickxi*) are more closely related to each other than the species of the San Cristobal-Española cluster (H. snodgrassi, H. junco, H. espanola). Within this San Cristobal-Espanola cluster, H. *espanola* often constitutes a separate cluster, which was further supported by a PCoA. These results show that, for the San Cristobal-Espanola cluster, intra island divergence of different ecotypes smaller than inter island divergence for the same ecotypes. Notwithstanding the fact that pairwise measurements of F_{st} between population pairs of Isabela and Santa Cruz are low in general, a similar, but less prominent, pattern emerged on these islands. Different ecotypes inhabiting the same island (H. *hendrickxi*, H. *galapagoensis*) tended to be genetically less differentiated compared to similar ecotypes populating different islands (H. *galapagoensis*). Hence, genetic differentiation increases with island age. This pattern is consistent with the age of the islands, for which San Cristobal and Espanola are the older islands and the Isabela-Santa Cruz cluster is comprised of the younger islands. The species of the older island cluster are more strongly diverged than the species on the younger island cluster. This might be caused by two factors: i) on the older islands, more time was available to diverge and ii) Isabela, Santiago and Santa Cruz are defined as the 'core' islands of the Galapagos. Generated paleogeographical reconstructions reveal that since 700 ka, changes in sea level have led to changes in connectivity. Especially when the sea level was low, this resulted in very high connectivity (Ali and Aitchison 2014), presumably leading to higher gene flow among these islands.

F_{st} values are twice as high for the galapagoensis clade dataset compared with the F_{st}-values of the complete dataset. This might be caused by divergent sites that are included in the galapagoensis dataset, but not present in H. jacquesbreli and H. albemarlensis, as these sites might have diverged too much to be aligned to the reference RAD tags. However, patterns remain consistent irrespective of which dataset is considered.

Patterns of genomic divergence

A large number of SNPs showed strong genetic divergence between similar ecotypes residing on different islands, and this pattern was observed in both low- and highland ecotypes. Remarkably, the

majority of SNPs characterized by a strong divergence between both lowland ecotypes, were exactly the same SNPs showing strong genetic differentiation between both highland ecotypes. Such large amount of shared, strongly differentiated genetic variation might indicate divergent selection between the islands San Cristobal and Santa Cruz (Seehausen et al. 2014). This corresponds with the morphological analysis of genital and non-genital traits (De Busschere et al. 2012) in which they found that the genital traits are probably under sexual selection and that these traits show strong similarity within islands. Hence, these observed outlier SNPs might be linked to sexual traits, either directly or indirectly through linkage disequilibrium of the SNP with a region under selection.

No SNPs with a high Fst value are shared between the San Cristobal (H. *junco* - H. *snodgrassi*) and the Santa Cruz (H. *hendrickxi* - H. *galapagoensis*) comparison. Although, 40 RAD tags on which the outlier SNPs are positioned are shared between San Cristobal and Santa Cruz. This result is unexpected, as the species belonging to the same ecotype showed a high phenotypical similarity (De Busschere et al. 2012). Many genes could be linked to the phenotypic traits of these ecotypes, with small allele frequency differences at many loci, which do not show a high F_st value. Another possibility is that the phenotypes have evolved through different genetic and developmental pathways (Arendt and Reznick 2008). This latter pattern has been observed for the evolution of flat wings in Crickets in Hawaii (Pascoal et al. 2014).

Phylogenetic relationship

Corroborating previous results, the phylogenetic relationship based on the multispecies coalescent revealed monophyletic clade for the 'top' and 'low' ecotypes species intra island. These results should be interpreted cautiously, as despite the node high probabilities, the Effective Sample Size for the likelihood and posterior probability were lower than recommended.

Superimposing the Maximum likelihood trees for the neutral and outlier loci revealed wide spread gene tree incongruence, possibly by different evolutionary histories due to incomplete lineage sorting and interspecific gene flow (Hou et al. 2015; Nater et al. 2015). Often, only neutral loci are used, as these loci are not under selection and are considered to resemble the divergence history. Although, episode of second contact or ongoing gene flow erases the initial signal of differentiation (Bierne, Gagnaire, and David 2013). Moreover, even if the previous factors aren't acting, gene trees can have a large variance just by chance (Nater et al. 2015).

The inability to find a clear pattern might also be due to methodological artefacts, like assembly, mapping, null alleles and filtering options. Missing data (positions that are unknown), can be generated during several steps of RAD sequencing. First, a mutation at the enzyme-cutting sites or a newly mutated enzyme-cutting site can cause the loss of a fragment, i.e. null allele, in some taxa included in a study (Hohenlohe et al. 2010; Huang and Knowles 2014; McCormack et al. 2013; Ree and Hipp 2015). The null alleles generated by this process are strongly linked to the genetic distance among the taxa. The data processing step can create missing data through the filtering on quality and coverage. Mapping to the reference tags also generates null alleles, mainly when more divergent taxa are mapped. H. *albemarlensis* and H. *jacquesbreli* have the highest percentage of missing data and H. *galapagoensis* and H. *hendrickxi* the lowest percentage, as these species are respectively the least and most closely related to the individual of the reference tags. Arnold et al. 2013 conducted research to assess the effect of this non-random sampling on evolutionary parameters. They concluded from their simulation studies that

loci with missing data gave inaccurate estimates of summary statistics (i.e. F_{st}) and may increase the rate of false positives in outlier analyses. Furthermore, a large part of true F_{st} outliers could be absent, as many true outliers have incomplete sampling (Arnold et al. 2013). Despite these issues, mapping against reference tags gives the best results compared with de novo analysis (Nadeau et al. 2014).

There is a strong need for studies that research the performance and limits of species-tree inference with RADseq and the development of a standard protocol (Ree and Hipp 2015). Nonetheless, using RAD data to infer phylogenies is very promising for future research.

Recommendations

For future research, we recommend to acquire a higher coverage for the assembly as this results in greater quality reference tags. We suggest two approaches to increase the coverage of the reads for the reference RADtags assembly, either sequencing more reads or pooling the reads of multiple individuals of the same species for assembly.

This study mainly focused on the divergence patterns of San Cristobal and Santa Cruz. For future research we would recommend including H. *espanola* and H. *galapagoensis* from Isabela in the analysis and observing whether an identical or different ecotype comparison results in identical or different outlier SNPs.

As gene flow can have a strong effect on radiations, it would be interesting to estimate the relative importance of incomplete lineage sorting versus gene flow (Hendrickx et al. 2015) by implementing models that estimate the rate of inter- and intra-island gene exchange (Hendrickx et al. 2015; Hey and Nielsen 2004).

7. Conclusion

There is a strong indication that parallel evolution has occurred within the *Hogna* species on the Galapagos, considering that the species cluster according geography and intra-island divergence increases with island age. There is no evidence found for a shared genetic background leading to the identical ecotypes. The underlying mechanisms of the parallel evolution might be (i) due to a different genomic architecture linked to the evolution of the ecotypes, or (ii) multiple genes with small differences in allele frequencies influence the phenotype. The inter-island differentiation analysis indicates that divergent selection is acting. This coincides with a previous study, which concluded that genital traits were more similar within islands and are probably under sexual selection. The gene trees of neutral and outlier loci revealed a high incongruence, which might be due to incomplete lineage sorting, interspecific gene flow or methodological artefacts. Future research is highly recommended to improve the reliability of the analysis and to further unravel the mechanisms influencing the parallel evolution within the *Hogna* species.

8. Summary

A remarkable observation on island archipelagos is that a single species may adaptively radiate into a set of different species. The presence of comparable environments on different islands may thereby result in in ecologically similar species, i.e. ecotypes. The evolution of each ecotype can occur once,

followed by colonization of the different islands, or can occur repeatedly on different islands. The latter pattern is called parallel evolution, defined as the repeated evolution of similar phenotypes. Traits that evolve in parallel can be the result of different genetic backgrounds evolving in different populations, or by repeated selection of the same allele that is present as standing genetic variation or is introduced by introgression from nearby islands. Advances in NGS made it possible to uncover thousands of markers across a genome and resulted in the emergence of new analytical approaches to study which mechanism underlie parallel evolution.

The *Hogna* radiation from the Galapagos constitute a perfect system to investigate which processes and patterns shape parallel evolution, as divergence into similar ecotypes occurred repeatedly within the archipelago. The first ecotype is called the 'top' ecotype and is found on the top of the islands. The second ecotype is called the 'low' ecotype and distributed along the drier coastal region. A previous study has shown that species belonging to the same ecotype, found on different islands, share a range of phenotypic similarities while similarity in genital traits is determined by geography.

We made use of Restriction-site Associated DNA sequencing (RADseq) to explore the genetic relationships between the species, and a set of 84 individuals, comprising all extant species, was subject to analysis.

Investigating the genetic relationship between the species by means of PCoA, Bayesian clustering and genome wide measures of differentiation were consistent and revealed that the species *H. jacquesbreli*, being the highland species from the young island Isabela and H. *albemarlensis*, being a generalist species occurring on all islands, are most strongly differentiated from the remaining species. This latter species group, further referred to as the 'galapagoensis clade' constitutes the species group wherein the repeated evolution of high- and lowland species is most apparent. Within this clade, species form two clusters according geography i.e. (i) the Isabela-Santa Cruz cluster with the highland species H. *galapagoensis* and the lowland species H. *hendrickxi* and (ii) the Española-San Cristóbal cluster with the lowland species H. *snodgrassi* and H. *espanola* and the highland species H. *junco*. The species on the older islands, Espanola and San Cristobal, show higher divergence in contrast to the species of the younger islands, Isabella-Santa Cruz. In general, degree of genetic differentiation between ecotypes within islands was very low (Fst between 0.109 and 0.036), indicating that despite their strong phenotypic divergence in a suite of traits, ecotypes from the same island are very closely related to each other.

Next we determined genomic differentiation to identify neutral and adaptive loci which can reveal genomic regions that show evidence of divergent selection. To identify candidate loci that potentially are associated with sites subjected to natural selection, we tested for the presence of loci which are significantly more differentiated, i.e. outlier loci. These were determined for the ecotypes intra-island and for each ecotype inter-islands. The first analysis detected a substantial number of outlier loci within islands (0.79 % on Santa Cruz and 1.46 % on San Cristóbal), and indicate genomic regions associated with divergent selection between the ecotypes within islands. Remarkably, these outlier loci appeared to be hardly shared between islands, which suggests that the repeated evolution of these ecotypes is the result of selection in different regions of the genome on different islands. Although this points in the direction of truly independent evolution, it cannot be excluded that this is the result of small differences in allele frequencies on similar loci that remained undetected by the outlier analysis.

We also detected a high number of shared outlier loci for the inter-island comparison, indicating divergent selection, which might be the result of divergent selection on sexual traits.

Reconstructing the phylogenetic relationship based on the multispecies coalescent showed clustering according geography. Superimposing the Maximum likelihood trees constructed for the neutral and outlier RAD tags revealed that a considerable amount of variation is present among the individual trees. This might either be due to incomplete lineage sorting, interspecific gene flow or an insufficient number of polymorphic sites in the individual RADtags to reliable reconstruct phylogenies from individual RADtags.

We can conclude that the results strongly indicate that parallel evolution effectively took place within this *Hogna* radiation, and that no clear signal was present that points to repeated selection of similar alleles on different islands. As this study constitutes a first attempt to infer the genomic basis behind this radiation, we further provide recommendations to improve the reliability of the analyses in future research.

9. Samenvatting

Een opmerkelijk fenomeen op eilandengroepen is dat een soort zich mogelijks adaptief evolueert in verschillende soorten. De aanwezigheid van vergelijkbare omgevingen op verschillende eilanden kan leiden naar soorten met een gelijkaardige ecologie, ecotypes genoemd. Deze evolutie kan eenmalig voorkomen, gevolgd door dispersie naar verschillende eilanden, of kan herhaaldelijk voorkomen op verschillende eilanden. Het laatste patroon wordt parallelle evolutie genoemd, gedefinieerd als het herhaaldelijke evolueren van dezelfde fenotypes. Dit kan het gevolg zijn van evolutie van dezelfde fenotypes via een verschillende genetische achtergrond, of de herhaaldelijke selectie op hetzelfde allel dat aanwezig is als staande genetische variatie of dat is geïntroduceerd via introgressie van nabije eilanden.

De vooruitgang in NGS maakt het mogelijk om duizenden merkers over het volledige genoom te detecteren en resulteert in de opkomst van nieuwe analytische aanpakken om het mechanisme van parallelle evolutie te bestuderen.

The *Hogna* radiatie van de Galapagoseilanden is het perfecte systeem om de processen en patronen die parallelle evolutie vormen te bestuderen, aangezien divergentie in gelijkaardige ecotypes verschillende malen heeft plaatsgevonden binnen de archipel. Het eerste ecotype wordt de 'top' soort genoemd en wordt op de toppen van de eilanden gevonden. Het tweede ecotype wordt de 'beneden' soort genoemd en komt voor in droge kusthabitats. Een voorgaande studie heeft aangetoond dat de soorten die behoren tot hetzelfde ecotype, maar op een verschillend eiland voorkomen, fenotypische sterk op elkaar lijken. De gelijkenissen voor genitale kenmerken worden daarentegen door geografie bepaald.

Het onderzoek van de genetische relatie tussen de soorten via een Principal Component Analysis, Bayesiaanse clustering en de mate van genoomwijde differentiatie waren consistent met elkaar. Ze toonden aan dat H. *jacquesbreli*, een 'top' soort voorkomende op het jongere eiland Isabella, en H. *albemarlensis*, de generalistische soort verspreid over de volledige archipel, sterker zijn gedifferentieerd van de overige soorten. Naar deze laatste soortengroep wordt verwezen als de 'galapagoensis' clade waarin de herhaaldelijke evolutie van 'top' en 'beneden' soorten het meest waarschijnlijk is. Binnen

deze clade ontstonden er twee clusters volgens geografie , namelijk de Isabella en Santa-Cruz cluster met de topsoort H. *galapagoensis* en de beneden *soort* H. *hendrickxi,* en de Española en San Cristobal cluster met de benedensoorten H. *snodgrassi* en H. *espanola* en de topsoort H. *junco*.

Vervolgens werd genomische differentiatie vastgesteld om neutrale en adaptive loci te identificeren. Deze genomische regio's kunnen een bewijs vormen voor divergente selectie. Om loci te identificeren die mogelijk geassocieerd zijn met posities onder natuurlijk selectie, hebben we getest op de aanwezigheid van loci die meer significant gedifferentieerd zijn, nl. outlier loci. Deze werden bepaald voor de ecotypes binnen elk eiland en voor dezelfde ecotypes tussen eilanden. De eerste analyse detecteerde een substantieel aantal outlier loci binnen eilanden (0.79% voor Santa Cruz en 1.46% voor San Cristobal) die mogelijk geassocieerd zijn met divergente selectie tussen de ecotypes binnen eilanden. Opmerkelijk is dat deze outlier loci niet gedeeld zijn tussen eilanden, wat suggereert dat de herhaaldelijke evolutie van deze ecoytypes het resultaat is van selectie op verschillende regio's in het genoom op verschillende eilanden. Ondanks het feit dat dit sterk wijst op onafhankelijk evolutie, kan het niet worden uitgesloten dat kleine verschillen in allel-frequenties van dezelfde loci ongedecteerd blijven door de outlier analyse.

Voor de vergelijking tussen eilanden werd een hoog aantal outliers gedetecteerd, wat wijst op divergente selectie. Dit zou het resultaat kunnen zijn van divergente selectie op seksuele kenmerken tussen eilanden.

De reconstructie van de fylogenetische relaties gebaseerd op coalescentie van meerdere loci toonde de clustering volgens geografie aan. Het samenleggen van de Maximum likelihood bomen van neutrale en adaptieve RADtags toonde aan dat er veel variatie aanwezig was tussen de individuele bomen. Dit kan worden veroorzaakt door incomplete lineage sorting, interspecifieke gene flow of een onvoldoende aantal polymorfische sites in de RADtags om betrouwbare fylogenetische reconstructies te maken.

We kunnen concluderen dat de resultaten sterk aanwijzen dat parallelle evolutie werkelijk plaats heeft gevonden binnen deze *Hogna*-radiatie en dat geen duidelijk signaal is gevonden van herhaaldelijke selectie op dezelfde allelen in verschillende eilanden. Deze studie draagt bij tot de eerste poging om de genomische achtergrond van deze radiatie onthullen en we raden aan om bij verdere studies de betrouwbaarheid van deze analyses te verbeteren.

10. References

Ali, Jason R., and Jonathan C. Aitchison. 2014. "Exploring the Combined Role of Eustasy and Oceanic Island Thermal Subsidence in Shaping Biodiversity on the Galápagos." *Journal of Biogeography* 41(7): 1227–41.

Andrew, S. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data."

Arendt, Jeff, and David Reznick. 2008. "Convergence and Parallelism Reconsidered: What Have We Learned about the Genetics of Adaptation?" *Trends in Ecology and Evolution* 23(1): 26–32.

Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies. 2013. "RADseq Underestimates Diversity and Introduces Genealogical Biases due to Nonrandom Haplotype Sampling." *Molecular Ecology* 22(11): 3179–90.

Baird, Nathan a. et al. 2008. "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers." *PLoS ONE* 3(10): 1–7.

Barrett, R. D H, and Dolph Schluter. 2008. "Adaptation from Standing Genetic Variation." *Trends in Ecology and Evolution* 23(1): 38–44.

Bierne, Nicolas, Pierre Alexandre Gagnaire, and Patrice David. 2013. "The Geography of Introgression in a Patchy Environment and the Thorn in the Side of Ecological Speciation." *Current Zoology* 59(1): 72–86.

Browning, Brian L., and Sharon R. Browning. 2007. "Efficient Multilocus Association Testing for Whole Genome Association Studies Using Localized Haplotype Clustering." *Genetic epidemiology* 31(8): 365–75.

Browning, Sharon R, and Brian L Browning. 2007. "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering." *American journal of human genetics* 81(5): 1084–97.

De Busschere, C. et al. 2010. "Parallel Habitat Specialization within the Wolf Spider Genus Hogna from the Galápagos." *Molecular Ecology* 19(18): 4029–45.

De Busschere, Charlotte et al. 2012. "Parallel Phenotypic Evolution in a Wolf Spider Radiation on Galápagos." *Biological Journal of the Linnean Society* 106(1): 123–36.

Butlin, Roger K. et al. 2014. "Parallel Evolution of Local Adaptation and Reproductive Isolation in the Face of Gene Flow." *Evolution* 68(4): 935–49.

Catchen, J. M. et al. 2011. "Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences." *G3journall* 1(3): 171–82.

Catchen, Julian et al. 2013. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22(11): 3124–40.

Danecek, Petr et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27(15): 2156–58.

Davey, John L., and Mark W. Blaxter. 2010. "RADseq: Next-Generation Population Genetics." *Briefings in Functional Genomics* 9(5-6): 416–23.

Davey, John W et al. 2011. "Genome-Wide Genetic Marker Discovery and Genotyping Using next-Generation Sequencing." *Nature reviews. Genetics* 12(7): 499–510.

DePristo, Mark A et al. 2011. "A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data." *Nature genetics* 43(5): 491–98.

Drummond, Alexei J., Simon YW Ho, Nic Rawlence, and Andrew Rambaut. 2007. "A Rough Guide to BEAST 1.4." *Edinburgh: …*: 1–41.

Earl, Dent A., and Bridgett M. vonHoldt. 2012. "STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method." *Conservation Genetics Resources* 4(2): 359–61.

Edelaar, Pim, Adam M. Siepielski, and Jean Clobert. 2008. "Matching Habitat Choice Causes Directed Gene Flow: A Neglected Dimension in Evolution and Ecology." *Evolution* 62(10): 2462–72.

Etter, Paul D. et al. 2011. "SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing (V Orgogozo, M V. Rockman, Eds,)." *Methods in molecular biology* 772: 157–78.

Foll, Matthieu, and Oscar Gaggiotti. 2008. "A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective." *Genetics* 180(2): 977–93.

Futuyma, J. Douglas. 2013. *Evolution*. Third edit. Sinauer Associates.

Garant, Dany, Samantha E. Forde, and Andrew P. Hendry. 2007. "The Multifarious Effects of Dispersal and Gene Flow on Contemporary Adaptation." *Functional Ecology* 21(3): 434–43.

Gillespie, Rosemary G. 2013. "Adaptive Radiation: Convergence and Non-Equilibrium." *Current Biology* 23(2): R71–74.

Gompel, Nicolas, and Benjamin Prud'homme. 2009. "The Causes of Repeated Genetic Evolution." *Developmental Biology* 332(1): 36–47.

Heled, J., and A. J. Drummond. 2010. "Bayesian Inference of Species Trees from Multilocus Data." *Molecular Biology and Evolution* 27(3): 570–80.

Hendrickx, Frederik et al. 2015. "Persistent Inter- and Intraspecific Gene Exchange within a Parallel Radiation of Caterpillar Hunter Beetles (Calosoma Sp.) from the Galápagos." *Molecular Ecology* 24(12): 3107–21.

Hey, Jody, and Rasmus Nielsen. 2004. "Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, with Applications to the Divergence of Drosophila Pseudoobscura and D. Persimilis." *Genetics* 167(2): 747–60.

Hohenlohe, Paul a. et al. 2010. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLoS Genetics* 6(2): 1–23.

Hou, Yan et al. 2015. "Thousands of RAD-Seq Loci Fully Resolve the Phylogeny of the Highly Disjunct Arctic-Alpine Genus Diapensia (Diapensiaceae)." *PLoS ONE* 10(10): 1–14.

Huang, Huateng, and L Lacey Knowles. 2014. "Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences." *Systematic biology* 0(0): 1–9.

Jakobsson, Mattias, and Noah A. Rosenberg. 2007. "CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure." *Bioinformatics* 23(14): 1801–6.

Lamichhaney, Sangeet et al. 2015. "Evolution of Darwin's Finches and Their Beaks Revealed by Genome Sequencing." *Nature* 518: 371–75.

Leibold, M. a. et al. 2004. "The Metacommunity Concept: A Framework for Multi-Scale Community Ecology." *Ecology Letters* 7(7): 601–13.

Losos, Jonathan B, and Robert E Ricklefs. 2009. "Adaptation and Diversification on Islands." *Nature* 457(7231): 830–36.

Lunter, Gerton, and Martin Goodson. 2011. "Stampy: A Statistical Algorithm for Sensitive and Fast Mapping of Illumina Sequence Reads." *Genome Research* 21(6): 936–39.

Martin, Simon H et al. 2013. "Genome-Wide Evidence for Speciation with Gene Flow in Heliconius Butterflies." *Genome research* 23: 1817–28.

McCormack, John E. et al. 2013. "Applications of next-Generation Sequencing to Phylogeography and Phylogenetics." *Molecular Phylogenetics and Evolution* 66(2): 526–38.

McKenna, Aaron et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome research* 20: 1297–1303.

Nadeau, Nicola J et al. 2014. "Population Genomics of Parallel Hybrid Zones in the Mimetic Butterflies, *H. Melpomene* and *H. Erato*." *Genome research* 24(8): 1316–33.

Nater, Alexander et al. 2015. "Resolving Evolutionary Relationships in Closely Related Species with Whole-Genome Sequencing Data." *Systematic Biology* 64(6): 1000–1017.

Niemiller, Matthew L., Benjamin M. Fitzpatrick, and Brian T. Miller. 2008. "Recent Divergence with Gene Flow in Tennessee Cave Salamanders (Plethodontidae: Gyrinophilus) Inferred from Gene Genealogies." *Molecular Ecology* 17(9): 2258–75.

Nosil, Patrick. 2008. "Speciation with Gene Flow Could Be Common." *Molecular ecology* 17(9): 2103–6.

Pascoal, Sonia et al. 2014. "Rapid Convergent Evolution in Wild Crickets." *Current Biology* 24(12): 1369–74.

Peakall, Rod, and Peter E. Smouse. 2012. "GenALEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-an Update." *Bioinformatics* 28(19): 2537–39.

Porras-Hurtado, Liliana et al. 2013. "An Overview of STRUCTURE: Applications, Parameter Settings, and Supporting Software." *Frontiers in Genetics* 4(MAY): 1–13.

Pritchard, J K, M Stephens, and P Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155(2): 945–59.

Rambaut, Andrew. 2006. "Figtree."

Rambaut, Andrew, and A. J. Drummond. 2004. "TRACER (MCMC Trace Analysis Tool)."

Ree, Richard H, and Andrew L Hipp. 2015. "Inferring Phylogenetic History from Restriction Site Associated DNA (RADseq)." *Next-Generation Sequencing in Plant Systematics*: 1–24.

Rosenberg, Noah A. 2004. "DISTRUCT: A Program for the Graphical Display of Population Structure." *Molecular Ecology Notes* 4(1): 137–38.

Rosenblum, Erica Bree, Christine E. Parent, and Erin E. Brandt. 2014. "The Molecular Basis of Phenotypic Convergence." *Annual Review of Ecology, Evolution, and Systematics* 45(1): 203–26.

Rousset, François. 2008. "GENEPOP'007: A Complete Re-Implementation of the GENEPOP Software for Windows and Linux." *Molecular Ecology Resources* 8(1): 103–6.

Schluter, Dolph. 2000. *The Ecology of Adapative Radiation*. Oxford: Oxford University Press.

Seehausen, Ole et al. 2014. "Genomics and the Origin of Species." *Nature reviews. Genetics* 15(3): 176–92.

Smadja, Carole M., and Roger K. Butlin. 2011. "A Framework for Comparing Processes of Speciation in the Presence of Gene Flow." *Molecular Ecology* 20(24): 5123–40.

Smouse, P E, and R Peakall. 1999. "Spatial Autocorrelation Analysis of Individual Multiallele and Multilocus Genetic Structure." *Heredity* 82 (Pt 5)(January): 561–73.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13.

Stern, David L. 2013. "The Genetic Causes of Convergent Evolution." *Nature reviews. Genetics* 14(11): 751–64.

Warren, Ben H. et al. 2015. "Islands as Model Systems in Ecology and Evolution : Prospects Fifty Years after MacArthur-Wilson." *Ecology Letters* 18: 200–217.

Zerbino, Daniel R., and Ewan Birney. 2008. "Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs." *Genome Research* 18(5): 821–29.

11. Appendix

Table of all individuals included in this research. Individuals of which DNA concentration was to low weren't included in the RAD libraries. For each individual is the following information given: species name; year, island and location of sampling; RAD library number and Sequencing platform.

ID	Species	Year	Island	Location	Library	Sequencing
H309_065	H. albemarlensis	2009	Santiago	Los Jaboncillos	-	-
H309_066	H. albemarlensis	2009	Santiago	Los Jaboncillos	-	-
H309_069	H. albemarlensis	2009	Isabela	Cerro Azul	Lib75	MiSeq
H309_074	H. albemarlensis	2009	Isabela	Cerro Azul	-	-
H309_101	H. albemarlensis	2009	Santa Cruz	Laguna Andreas	-	-
H309_102	H. albemarlensis	2009	Santa Cruz	Laguna Andreas	-	-
H309_103	H. albemarlensis	2009	Santa Cruz	Laguna Andreas	-	-
H309_104	H. albemarlensis	2009	Isabela	Alcedo	Lib75	MiSeq
H309_105	H. albemarlensis	2009	Isabela	Alcedo	-	-
H309_106	H. albemarlensis	2009	Isabela	Alcedo	-	-
H509_046	H. albemarlensis	2009	San Cristóbal	Puerto Grande	Lib71	HiSeq
H709_001	H. espanola	2009	Española	Bahia Manzanilla	Lib71	HiSeq
H709_002	H. espanola	2009	Española	Bahia Manzanilla	Lib71	HiSeq
H709_003	H. espanola	2009	Española	Bahia Manzanilla	Lib72	HiSeq
H709_004	H. espanola	2009	Española	Bahia Manzanilla	Lib72	HiSeq
H709_019	H. espanola	2009	Española	Bahia Manzanilla	Lib73	HiSeq
H709_032	H. espanola	2009	Española	Bahia Manzanilla	Lib73	HiSeq
H709_034	H. espanola	2009	Española	Bahia Manzanilla	Lib74	HiSeq
H709_035	H. espanola	2009	Española	Bahia Manzanilla	Lib74	HiSeq
H709_036	H. espanola	2009	Española	Bahia Manzanilla	Lib72	HiSeq
H709_037	H. espanola	2009	Española	Bahia Manzanilla	Lib75	MiSeq
H409_001	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib71	HiSeq
H409_002	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib71	HiSeq
H409_003	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib72	HiSeq

H409_004	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib72	HiSeq
H409_005	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib73	HiSeq
H409_006	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib73	HiSeq
H409_007	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib74	HiSeq
H409_008	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib74	HiSeq
H409_009	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib76	HiSeq
H409_010	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib75	MiSeq
H409_011	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib71	HiSeq
H409_101	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	Lib73	HiSeq
H409_102	H. galapagoensis	2009	Santa Cruz	Between Media Luna and Cerro Crocker	-	-
H414_001	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_002	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_003	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_004	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_005	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_006	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_007	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H414_008	H. galapagoensis	2014	Isabela	Alcedo	Lib75	MiSeq
H609_003	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib71	HiSeq
H609_004	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib71	HiSeq
H609_005	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib72	HiSeq
H609_011	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib72	HiSeq
H609_012	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib73	HiSeq

H609_014	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib73	HiSeq
H609_015	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib74	HiSeq
H609_021	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib74	HiSeq
H609_022	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib75	MiSeq
H609_023	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib72	HiSeq
H609_024	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib75	MiSeq
H609_025	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib75	MiSeq
H609_026	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib74	HiSeq
H609_027	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib75	MiSeq
H609_028	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib74	HiSeq
H609_034	H. hendrickxi	2009	Santa Cruz	Las Palmas	Lib76	HiSeq
H109_001	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib74	HiSeq
H109_002	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib71	HiSeq
H109_003	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib72	HiSeq
H109_004	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib72	HiSeq
H109_005	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib73	HiSeq
H109_006	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib73	HiSeq
H109_007	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib74	HiSeq
H109_008	H. jacquesbreli	2009	Isabela	Top Sierra Negra	Lib74	HiSeq
H209_001	H. junco	2009	San Cristóbal	El Junco	Lib71	HiSeq
H209_002	H. junco	2009	San Cristóbal	El Junco	Lib71	HiSeq
H209_003	H. junco	2009	San Cristóbal	El Junco	Lib71	HiSeq
H209_004	H. junco	2009	San Cristóbal	El Junco	Lib72	HiSeq
H209_005	H. junco	2009	San Cristóbal	El Junco	Lib72	HiSeq
H209_006	H. junco	2009	San Cristóbal	El Junco	Lib72	HiSeq
H209_008	H. junco	2009	San Cristóbal	El Junco	Lib73	HiSeq
H209_009	H. junco	2009	San Cristóbal	El Junco	Lib73	HiSeq
H209_010	H. junco	2009	San Cristóbal	El Junco	Lib73	HiSeq
H209_011	H. junco	2009	San Cristóbal	El Junco	Lib74	HiSeq
H209_012	H. junco	2009	San Cristóbal	El Junco	Lib76	HiSeq

H209_013	H. junco	2009	San Cristóbal	El Junco	Lib74	HiSeq
H209_016	H. junco	2009	San Cristóbal	El Junco	Lib71	HiSeq
H209_101	H. junco	2009	San Cristóbal	El Junco	-	-
H209_102	H. junco	2009	San Cristóbal	El Junco	-	-
H209_103	H. junco	2009	San Cristóbal	El Junco	Lib73	HiSeq
H209_104	H. junco	2009	San Cristóbal	El Junco	-	-
H509_031	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib71	HiSeq
H509_043	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib71	HiSeq
H509_047	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib72	HiSeq
H509_048	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib72	HiSeq
H509_049	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib72	HiSeq
H509_051	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib73	HiSeq
H509_052	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib73	HiSeq
H509_053	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib73	HiSeq
H509_054	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib74	HiSeq
H509_055	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib76	HiSeq
H509_056	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib74	HiSeq
H509_058	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib71	HiSeq
H509_059	H. snodgrassi	2009	San Cristóbal	Puerto Grande	Lib74	HiSeq