Distinct biogeographic origins of androgenetic *Corbicula* lineages followed by genetic captures

Etoundi E.[1], Marescaux J. [1,2], Vastrade M. [1], Debortoli N. [1,2], Hedtke S.M.[3], Pigneur L.-M. [4], Virgo J.[1], Flot J.-F.[5,6] and Van Doninck K.[1]

[1] Laboratory of Evolutionary Genetics and Ecology; Research Unit in Environmental and Evolutionary Biology; Institute of Life, Earth and Environment (ILEE); University of Namur, 5000 Namur, Belgium

[2] e-biom SA, 5/7 Rue Godefroid, 5000 Namur, Belgium

[3] Department of Animal, Plant and Soil Sciences, School of Life Sciences, La Trobe University, Bundoora, VIC 3086, Australia

[4] Université de Liège, Laboratoire de génétique de la conservation
Chemin de la vallée 4, 4000 Liège, Belgium

[5] Université Libre de Bruxelles (ULB), Evolutionary Biology & Ecology, C.P. 160/12, Avenue F.D. Roosevelt 50, 1050 Brussels, Belgium

[6] Interuniversity Institute of Bioinformatics in Brussels – (IB)[2], Brussels, Belgium

Corresponding author: Karine Van Doninck
Fax: +3281724362
E-mail: karine.vandoninck@unamur.be

Abstract

Corbicula clams were introduced during the 20th century into America and Europe, where they became notoriously successful invaders with a widespread, global distribution. Their ability to reproduce clonally through androgenesis ("all-male asexuality") has been determinant in their invasive success, with only four invasive clonal lineages detected across Europe and America, one of which is very abundant and widespread on both continents. Due to their "all-male asexuality" and egg parasitism between distinct lineages, the evolutionary and geographic origins of the invasive androgenetic lineages have been challenging to identify. We analyzed here the patterns of allele sharing for different molecular markers among *Corbicula* individuals collected worldwide. We identify three distinct genetic pools containing androgenetic *Corbicula* lineages. While one sexual *Corbicula* species forms a distinct fourth genetic pool, the other sexual lineages cluster with the androgenetic ones based on shared alleles. One genetic pool contains most androgenetic lineages and sexual *C. sandai* from Lake Biwa in Japan, pointing to this lake as a likely origin of androgenetic *Corbicula* lineages. Although three distinct biogeographic origins of *Corbicula* androgenetic lineages have been identified, their recent radiation and cross-lineage genetic mixing hamper classical species delimitation within this clam genus.

Keywords: androgenesis, asexuality, hybridization, invasion, allele sharing

Introduction

Androgenesis (also known as "all-male asexuality") is an unusual reproductive mode in which the offspring inherits their entire nuclear genome from the gametes of their father only. Either maternal nuclear chromosomes are discarded from eggs, or oocytes are already 'non-nucleated' prior to fertilization (reviewed in Schwander & Oldroyd, 2016). The clam genus *Corbicula* is the only metazoan clade in which obligate androgenesis has been identified, with multiple lineages globally distributed and for which the underlying cytological mechanism is well described (Komaru et al. 2000; Ishibashi et al. 2002, 2003; Ishibashi & Komaru, 2006; Hotta & Komaru, 2018). Androgenetic *Corbicula* clams are hermaphroditic and produce offspring after fertilization of a congeneric oocyte by a biflagellate, unreduced sperm or through self-fertilization (Kraemer *et al.*, 1986). The oocyte then undergoes an abnormal meiosis in which the rotation of the meiotic axis does not occur and all the maternal nuclear chromosomes are discarded from the egg as two polar bodies (Komaru *et al.* 2000; Hotta & Komaru, 2018). The zygote therefore inherits the mitochondria from the oocyte and the nuclear genome from the sperm (reviewed in Pigneur *et al.*, 2012). Besides the hermaphroditic androgenetic lineages, sexual lineages are also found within the genus *Corbicula* and can occur in sympatry with androgens. Sexual *Corbicula* lineages appear to be dioecious, producing reduced, uniflagellate sperm prior to fertilization (Glaubrecht *et al.*, 2006).

Although androgenesis can be considered sexual from a functional and ecological perspective (eggs and sperm are required and fertilization does occur, albeit without karyogamy), androgenesis is asexual from a genetic point of view since the entire paternal nuclear genome is transmitted clonally. However, the retention of sexual features (*i.e.*, meiosis and fertilization) in *Corbicula* clams results in egg parasitism with mitochondrial and/or nuclear capture (Hedtke & Hillis, 2011). Indeed, when androgenesis occurs between distinct genetic lineages of *Corbicula*, the nuclear genome of one lineage becomes associated with the mitochondria of another lineage, resulting in egg parasitism (characterized by "cytonuclear mismatch") and increasing the diversity of mitochondrial sequences within a lineage (Figure 1a) (Lee *et al.*, 2005; Hedtke *et al.*, 2008; Pigneur *et al.*, 2012). Moreover, in some cases (never quantified), maternal nuclear chromosomes may be partially retained and both maternal and paternal nuclear chromosomes inherited by the offspring resulting in an increased ploidy (Figure 1b) (Komaru *et al.*, 2006). Such nuclear capture events, if they occur among *Corbicula* lineages, may generate nuclear diversity in otherwise asexually reproducing lineages (Hedtke *et al*., 2011) and may have contributed to polyploidy and genetic divergence among androgenetic *Corbicula* lineages (Pigneur *et al.*, 2014; Schwander & Oldroyd, 2016).

Altough *Corbicula* clams are native to Africa, Asia, Australia and the Middle East (Araujo *et al.* 1993), sexual *Corbicula* lineages have very restricted geographic distributions: *Corbicula sandai* is endemic to Lake Biwa in

Japan, *Corbicula japonica* is restricted to East Asian brackish water and a few sexual *Corbicula* lineages are restricted to Indonesia (Glaubrecht *et al.* 2006; Yamada *et al.*, 2014). By contrast, androgenetic *Corbicula* have a widespread distribution across Asia and are successful invasive lineages in Europe, North America and South America where they have been introduced during the 20th century (Lee *et al.*, 2005; Hedtke *et al.*, 2008; Pigneur *et al.*, 2014; Gomes *et al.*, 2016). The ability of several *Corbicula* lineages to reproduce through androgenesis and to self-fertilize (Kraemer *et al.*, 1986) has likely facilitated such a large and rapid spread. However, only few distinct invasive *Corbicula* lineages have been detected in the invasive range, all of them androgenetic. For instance, three invasive morphotypes have been reported from America: form A and form B from North America and form C from South America (Siripattrawan *et al.* 2000; Lee *et al.* 2005). In Europe, three distinct morphotypes have also been described: form R, form S and form Rlc (Renard *et al.* 2000, Pfenninger *et al.* 2002, Marescaux *et al.* 2010). These invasive forms appear to exhibit virtually no genetic diversity over their invasive range at all the nuclear and mitochondrial markers tested (Hedtke *et al.*, 2008; Pigneur *et al.*, 2014; Gomes *et al.*, 2016), and seem to possess only four, unique mitochondrial COI haplotypes (Pigneur *et al.*, 2012, 2014): haplotype FW5 (form A/R), haplotype FW17 (form C/S), haplotype FW1 (form B) and haplotype FW4 (form Rlc). As the latter two only differ at one SNP in COI, they were referred to as RlcB by Peñarrubia *et al.*, 2017; however, we still distinguish these two forms in our present work because of their very distinct microsatellite genotypes and their different morphologies (Pigneur *et al.* 2014). When distinct invasive lineages are found in sympatry, cytonuclear mismatches and hybrids are often detected (Lee *et al.*, 2005; Hedtke *et al.*, 2008; Pigneur *et al.*, 2012, Peñarrubia *et al.*, 2017; Bespalaya *et al.*, 2018). The fourth form ("form D") described in North America (Tiemann *et al.* 2017) may also be a hybrid between form A/R and form B, as the morphotype is intermediate between these two forms, the COI being from form A/R (Siripattrawan *et al.* 2000; Pfenninger *et al.* 2002; Lee *et al.* 2005; Pigneur *et al.* 2011a) and the 28S being found in form B (Hedtke *et al.* 2008) as well as in Asia (Komaru *et al.* 2013).

This low diversity found in invasive *Corbicula* lineages is thought to result from the recent introduction of few individuals (in the 1920's in America, in the 1980's in Europe) and their rapid expansion through androgenetic reproduction (Mouthon, 1981, McMahon, 1892, Pigneur *et al.*, 2014; Gomes *et al.*, 2016; Bespalaya *et al.*, 2018). In contrast, androgenetic *Corbicula* from Lake Biwa in Japan (native area) are characterized by a high genetic diversity, comparable to that of sexual *Corbicula* lineages (Pigneur *et al.*, 2014). This could be the consequence of genetic captures by androgenetic lineages in the native range where many lineages co-occur, or indicate multiple transitions to asexuality from sexually reproducing populations (*e.g.* in ostracods, Bode *et al.*, 2010; in *Timema* stick insect, Schwander & Crespi, 2009; van der Kooi & Schwander, 2014), or both.

Studies investigating the genetic relationships within the genus *Corbicula* using mitochondrial and/or nuclear phylogenies found that androgenetic *Corbicula* lineages are non-monophyletic (Hedtke *et al.*, 2011; Pigneur *et al.* 2011a, 2012, 2014; Peñarrubia *et al.*, 2017). These observations could be explained either by a single origin of androgenetic *Corbicula* lineages, followed by independent nuclear capture events between sympatric lineages or by multiple, independent origins from diverse sexual lineages similarly followed by nuclear captures. Hedtke *et al*. (2011) proposed, based on phylogenetic reconstructions using nuclear markers, a common origin for the invasive androgenetic *Corbicula* lineages likely from a *C. sandai* ancestor based on the sharing of a common allele. The same authors suggested a second origin of androgenesis for the invasive form C/S since this lineage appeared to group with *Corbicula moltkiana* from Indonesia across their phylogenies. However, in the COI phylogenies of Pigneur *et al.*, form C/S clustered with *C. australis* (2011a) or with *C. fluminalis africana* (2014). These phylogenetic discordances due to egg parasitism and hybridization among *Corbicula* species, have until now hampered the investigation of phylogenetic relationships between androgenetic and sexual *Corbicula* lineages (see Hedtke & Hillis 2011; Hedtke *et al.* 2011; Pigneur *et al.*, 2012, 2014). Moreover, the origin of androgenetic lineages is not resolved yet.

In the present study, the relationship among *Corbicula* lineages with different reproductive modes was investigated using haplowebs (Flot *et al.*, 2010), an allele sharing-based approach to species delimitation that can delineate species regardless of their monophyly or lack thereof (Flot 2015). Haplowebs delineate fields for recombination (FFRs, Doyle 1995), *i.e.* groups of individuals that share a common allele pool (AP) distinct of those of other such groups. This method not only enabled us to study genetic relationships in *Corbicula* but also to pinpoint the origin of specific alleles and the genetic captures that have occurred between distinct lineages. The specimens used in this study were obtained from the worldwide distribution of *Corbicula*, focusing our sampling efforts on regions where there are multiple morphologically and genetically diverse *Corbicula* lineages, including sexual ones (*C. sandai* from Lake Biwa, estuarine *C. japonica* in Japan, and *C. loehensis*, *C. possoensis*, *C. matannensis* and *C. moltkiana* from Indonesia). Moreover, we also sampled co-occurring sexual and androgenetic *Corbicula* individuals, which is a critical source of information since sympatric populations including sexuals are putative hotspots for the origin of new asexual lineages (Simon *et al.*, 2003).

Material and methods

Specimen collection and characterization of reproductive mode.

A total of 447 individuals from 48 distinct localities were collected across the geographic range of *Corbicula*, including both native and invasive areas (see Table S1) and preserved in 96% ethanol. Native *Corbicula* populations are defined here based on the sampled geographic localities. However, since invasive *Corbicula* lineages encompass a wide geographic distribution both in Europe and America, they are referred to as the four invasive forms (A/R, B, Rlc and C/S) previously described (Pigneur *et al.*, 2014).

The reproductive mode of invasive individuals was determined based on spermatozoa morphology (as in Pigneur *et al.*, 2014) under the assumption that androgenetic individuals produce biflagellate sperm while sexual clams produce uniflagellate sperm (Komaru & Konishi, 1999). Owing to the very little genetic diversity observed within invasive lineages in previous studies (Lee *et al.*, 2005; Hedtke *et al.*, 2008; Pigneur *et al.*, 2011a, 2014), we analyzed less than 10 individuals from each invasive form for further genetic analyses.

For samples collected in *Corbicula*'s native range, the lineage identity was based on mitochondrial haplotype or on the literature. The reproductive mode was inferred from the sperm morphology when possible, from the literature available for the sampled population (lineage assignment and/or reproductive morphology) or, if these characteristics were indeterminate, classified as "undetermined". We successfully retrieved sequences for at least one marker in a total of 348 individuals (Table S1).

Genetic analyses.

DNA isolation, markers amplification and sequencing.

We extracted DNA either from the adductor muscles or from the foot using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's protocol. We sequenced four different markers: (1) a 545-bp fragment of the mitochondrial cytochrome *c* oxidase (COI) gene; (2) a 350-bp fragment of the nuclear 28S gene (28S); (3) a 623-bp fragment containing the third intron of the nuclear α-amylase gene (*amy*), and (4) a 341-bp fragment of the hypervariable putative intron of the α subunit of the adenosine triphosphate synthase (*atps*). PCR reactions for the four markers were carried out in a mix comprising 1-4 µl DNA, 1X Taq buffer (Promega), 0.2 mM nucleotides, 0.3-0.5 µM of each primer and 0.1 unit of *Taq* polymerase (Promega) in a total volume of 25 µl. The PCR conditions and primers used for each marker are described in Supporting information 1. PCR products were purified and sequenced on an automated sequencer at Genoscreen (Lille, France) or Beckman Coulter Genomics (Bishop's Strotford, United Kingdom).

Sequence editing, alignment and allele-sharing approach

COI sequences were manually edited using BioEdit (Hall 1999) and tested for sample cross-contamination using AutoConTAMPR (https://github.com/jnarayan81/autoConTAMPR). Among the 182 individuals sharing haplotype(s) between different locations, 150 (82.4%) were included on this analysis, the COI sequences being not available for the remaining 32 individuals. No contamination was detected among these tested shared haplotypes.

The nuclear alleles were retrieved from the forward and reverse chromatograms of each individual as described in Fontaneto *et al.* (2015), including chromatograms from 24 individuals from Hedtke (2009) and Hedtke *et al*. (2011). These sequences were assembled and cleaned using Sequencher 4 (Gene Codes). For most individuals, direct sequencing of nuclear genes yielded chromatograms with double peaks as expected for heterozygotes. When the sequences contained numerous double peaks over the entire chromatograms, thereby indicating length-variant heterozygotes, the two alleles were inferred following the procedure of Flot *et al.* (2006) implemented in CHAMPURU 1.0 (Flot 2007). Individuals presenting chromatograms with only one double peak were resolved 'by eye', whereas those with a few double peaks, as expected for heterozygotes with alleles of identical lengths, were resolved statistically using SeqPHASE (Flot 2010) and PHASE (Stephens *et al.*, 2001) by taking the known alleles of homozygotes and length-variant heterozygotes into account. Only allele pairs with a PHASE posterior probability ≥ 90% were retained in the analysis. The chromatograms from Hedtke (2009) and Hedtke *et al*. (2011) represented three to seven cloned PCR products per individual and were not expected to have a double peak; single mutations found in only one clone were not retained as they could be the result of PCR error rather than biological signal.

The absence of chimeric sequences in our genetic dataset was confirmed by using *Automated Chimaera* tool (default parameters) in RDP4 (Martin *et al.*, 2015), a program designed to detect DNA recombination patterns (data not shown).

Since androgenetic *Corbicula* clams are predominantly triploid (*e.g.* Lee *et al.*, 2005; Hedtke *et al.*, 2008), up to three alleles per nuclear locus were expected. However, examination of the chromatograms did not reveal triple peaks. The same pattern was observed in previous studies using microsatellite data (except for one marker, Pigneur *et al.* 2011b) and in the distribution of sequences from cloned amplicons (Hedtke *et al.*, 2011). It is nevertheless possible that the procedure used here underestimated allelic diversity if null alleles failed to amplify during the PCR step. To validate our approach, we sequenced some American (2), European (2) and Malagasy (1) individuals

using both cloning (done twice) and direct sequencing of PCR products: the consensus sequences obtained from cloning PCR products were identical to the sequences obtained from direct sequencing (data not shown).

All sequences retrieved within our study or provided by Shannon Hedtke (Hedtke *et al.* 2011) were compared to the Genbank dataset and new sequences were deposited on Genbank (accession numbers listed in Table S2).

Allele sequences were aligned using the online service of MAFFT version 7 (Katoh *et al*. 2017) with default parameters. Indels were treated as a 5th character state in downstream analyses. Sequences found only in the native range of *Corbicula* (Asia, Africa and Australia) were defined as "native alleles", whereas sequences present in the invasive region were called "invasive alleles" regardless of whether they were also found in the native region.

We investigated allele sharing using the haploweb approach (Doyle, 1995; Flot *et al.*, 2010) (Figure S1). The online tool HaplowebMaker (https://eeg-ebe.github.io/HaplowebMaker/) was used to build the raw haplowebs, which are haplotype networks on which curves are added connecting haplotypes found co-occurring in heterozygous individuals. A group of haplotypes linked together by heterozygotes forms an exclusive allele pool (AP), the corresponding groups of individuals is called a field for recombination (FFR) (akin to species, Doyle 1995). The haplowebs were edited manually for visualization and uniformity purposes: each FFR was represented in a different color. The numerous singletons (alleles encountered only once in the whole dataset) were not considered in the analysis since they are not informative about allele-sharing. For this, we used the "Remove singletons" option implemented on the program, and default settings for the other parameters.

Based on the allele-sharing information per marker provided by HaplowebMaker, we did build a "conspecificity matrix" as in Debortoli *et al.* (2016) using the newly developed online CoMa tool (https://eeg-ebe.github.io/CoMa/) with no calculation on heterospecific pairs (option 1), *i.e.* no value was given to absence of sharing. In this matrix, the conspecificity score for each pair of individuals is the number of markers for which these individuals belong to the same FFR. As the COI marker is haploid, no FFR could be defined for this marker; the score was thus based on whether individuals share the same haplotype or not (score 0 or 1). After computing the matrix, rows and columns were reordered to maximize the scores along the diagonal using the hierarchical clustering method implemented in the R package ''heatmap3'' (Zhao *et al.* 2014). FFRs appear as blocks along the diagonal with high conspecificity scores within the blocks (and lower scores among them except when hybrids are present).

Alleles shared between *Corbicula* lineages were also summarized and visualized in a circular plot using the command-line version of Circos (Krzywinski *et al.*, 2009). All individuals sampled were included in the Circos except *C. australis* (ZMB106607) from Australia and *C. fluminalis cunnigtoni* (ZMB170399) from Congo because

these lineages yielded sequences for only one (28S) out of the four markers. Thirteen distinct groups were manually defined in the Circos plot to highlight the allele-sharing patterns for all four markers (COI, 28S, *amy* and *atps*). The four invasive forms were considered as four distinct groups, as well as the brackish water sexual lineage *C. japonica*. Native populations were defined based on the sampled geographic localities (Japan, Lake Biwa, Hawaii, Madagascar, Indonesia, Vietnam and South Africa). Due to the low number of individuals sampled, we combined the few individuals from Korea, Taiwan, Thailand and Philippines with those from China, all together in a group named "Asia".

Genetic diversity

Nucleotide diversity and haplotype diversity (π and Hd, respectively) were calculated as the percentage of variable sites between 2 sequences (*i.e.*, uncorrected pairwise distance) using DNAsp (Librado & Rozas 2009). Gaps were however not considered, underestimating the diversity among COI haplotypes and nuclear alleles.

Results

Genetic clustering of Corbicula *lineages*

We successfully sequenced 252 COI fragments comprising 58 distinct haplotypes that differed at 111 polymorphic sites (Table S1). For 28S, sequences from 274 individuals were obtained (136 alleles that differed at 97 polymorphic sites). We obtained *amy* sequences from 230 individuals (90 alleles at 195 polymorphic sites). Finally, we amplified *atps* sequences from 239 individuals, which grouped into 189 alleles with 199 polymorphic sites (Table S1).

A relatively high haplotype diversity (Hd) but a low nucleotide diversity (π) was found in *Corbicula* (Table 1), indicating many low-frequency, genetically similar alleles. Numerous singletons were observed for each nuclear gene (97 allele sequences out of 136 for 28S, 52 out of 90 for *amy* and 164 out of 189 for *atps*). Finally, for all three nuclear markers, most *Corbicula* individuals were heterozygous: based on 28S and *amy*, only 15% and 18% of all tested individuals respectively were homozygous (41/274 and 42/230) while 36% (85/239) were homozygous for *atps*.

Based on the exclusive allele pools retrieved by the haploweb analysis, only four FFRs were delineated in *Corbicula* for all three nuclear markers on all tested individuals from a worldwide sampling, including sexual and androgenetic lineages (Fig. 2). The conspecificity matrix combining data from all four markers (28S, *amy*, *atps* and COI) and excluding only the individuals for which no shared haplotype was retrieved (n=316), also found four main blocks corresponding to the four FFRs of the haploweb (Fig. S3).

FFR1 represents the brackish sexual lineage *C. japonica* that had no shared alleles with any other lineage and therefore clustered separately for all three nuclear markers (Fig.2, black clusters). This was confirmed by the mitochondrial COI network (Fig. 2d) and by the conspecifity matrix (Fig. S3). FFR2 includes the androgenetic *Corbicula* lineages from Vietnam clustering separately for the three nuclear markers tested (Fig.2, pink clusters). The single Korean individual shared a 28S allele (Vt10a; Table S1) with the Vietnamese individuals and therefore clusters with FFR2 only in the 28S haploweb (Figure 2a). The clustering of androgenetic *Corbicula* lineages from Vietnam with the single individual from Korea was also found in the COI network (Fig. 2d). Single individuals from the putative sexual Indonesian lineages, *C. moltkiana* and *C. tobiae*, also cluster with androgenetic Vietnamese individuals in the *amy* haploweb and in the conspecificity matrix since they shared the α-amylase allele Vt1a; the other tested genes of these Indonesian lineages contained singletons (Table S1; Fig. 2b; Fig. S3). FFR3 contains the invasive androgenetic form C/S from Europe and South America and *C. fluminalis africana* from South Africa for both nuclear genes *amy* and *atps* (sharing exclusively the *amy* allele C1a and the *atps* allele U3b, Fig.2b, c, orange clusters; Table S1). In the 28S haploweb, only four individuals of form C/S clustered separately in FFR3 (Fig. 2a). The other C/S individuals and *C. fluminalis africana* clustered within FFR4 (Fig. 2a) because these C/S individuals shared 28S allele C2a with three individuals (both US and Europe) of form A/R from FFR4 (ccc7, xx11 and ggg2; Table S1) and another allele C2b with all *C. fluminalis africana* individuals. When removing these three form A/R individuals from the 28S haploweb analysis, a new FFR appeared containing only the invasive androgenetic form C/S and *C. fluminalis africana* (FFR3', green, Figure S2). This highlights an apparent hybridization event between forms A/R and C/S. This separate clustering of *Corbicula* lineages C/S and *C. fluminalis africana* was also found in the conspecificity matrix (Fig. S3) and confirmed by the mitochondrial COI network (Fig. 2d). Finally, the largest FFR4 in the haploweb analysis included, for all the markers, the sexual lineage *C. sandai* from Lake Biwa, as well as androgenetic individuals across the entire geographic distribution of the genus *Corbicula*, notably China, Taiwan, Japan including Lake Biwa, Hawaii, the Philippines, Madagascar, Congo and three out of the four invasive forms (forms A/R, B and Rlc (Fig. 2, purple clusters). This is because a few alleles found in the invasive forms A/R, B and Rlc (AA10a and AB15b for 28S, AA10b and DB23b for *amy* and the AB12a *atps* allele) are widespread, being found in these distinct locations in Asia and Africa (Table S1). However, no cluster assignment for *atps* could be given to the invasive form Rlc or to the Indonesian individuals, as their alleles are all singletons, or to *C. sandai*, as we were unable to amplify its *atps* marker. The fourth block in the conspecificity matrix corresponds to this FFR4. The red zones outside the four blocks in the matrix (Fig. S3) represent allele sharing signatures between FFRs: zone A is representing individuals with shared 28S allele

between FFR3 and FFR4 as explained previously, zone B represents individual 266690 from Korea, sharing *amy* and *atps* alleles with *C. sp.* form A/R from FFR4 and the 28S allele Vt10a from FFR2.

In the haploweb approach, the minor FFRs composed of only one or two isolated individuals were not considered further as they did not give information about allele sharing.

***Allelic distribution in the genus* Corbicula**

The Circos plot is a graphical visualization of the alleles shared between *Corbicula* individuals. For visual purpose, only shared alleles with the four invasive forms are colored (Fig. 3). Only the sexual *C. japonica* clustered in a separate group (FFR1) in the Circos plot, highlighting the absence of allele sharing. Within and between the other groups, alleles are shared among taxonomically distinct *Corbicula* populations. Notable, the androgenetic Vietnamese individuals (FFR2, Fig. 2) have a shared allele with sexual Indonesian lineages as well as with one Korean individual (Fig. 3, Table S1). *C.* sp. form C/S shared alleles with *C. fluminalis africana* from South Africa, but a few alleles of invasive form C/S are also retrieved in the invasive form A/R (28S allele C2a) and in Lake Biwa individuals from FFR4 (*amy* allele U4b) (Fig. 3, Table S1). The other three invasive forms A/R, Rlc and B shared alleles with *Corbicula* individuals from 18 different locations from Asia and Africa (Table S1), including the sexual lineage *C. sandai* (Fig. 3).

Alleles and COI haplotypes commonly found in the invaded range (28S AA10a, AA10b, AB11a, AB15b, yy12a, ggg2a; *amy* AA10b, Db23b; *atps* ggg2a, AB12a, yy12b; COI FW5 and FW1) were present in at least two locations within the native range (Table S1) where these loci were generally heterozygous for two nuclear markers, associated with an unshared native allele. Indeed 27.5% and 29.7% native individuals are heterozygous with a widespread “invasive” allele and a unique native allele for 28S and *amy* respectively. The alleles found only in the native range of *Corbicula* (both in sexual and androgenetic lineages), had very narrow geographic distributions, restricted to one sampling location (Table S1).

Some of these widespread “invasive” alleles were also shared among the different invasive lineages: for the 28S gene, allele AA10a is shared by forms A/R and Rlc and is also found in sexual *C. sandai;* for the *amy* gene, forms A/R and B share the widespread allele AA10b, also found in sexual *C. sandai* (as observed by Hedtke et al. (2011)); the widespread *amy* allele Db23b was found in form Rlc and Japanese populations (except Lake Biwa) while form Rlc and B shared *amy* allele Db22b (also found in Lake Biwa) (Fig. 3; Table S1).

The Japanese sampling location Seta River-Lake Biwa, which included the sexual *C. sandai*, had a relatively high proportion of unique native alleles (12/15 haplotypes for COI, 52/54 for 28S, 26/31 for *amy* and 21/22 for *atps*; Table S1), but also encompassed alleles from the four invasive lineages (Fig. 3). In the other Japanese

locations tested (except where *C. japonica* is found), alleles from the three invasive forms A/R, Rlc and B were found (Fig. 3).

Discussion

Within the clam genus *Corbicula*, the sexual and androgenetic lineages are characterized by different reproductive, morphological and ecological features which have been used to delimit species (Glaubrecht *et al.* 2003, Korniushin 2004, Lee *et al.* 2005, Rintelen & Glaubrecht, 2006, Marescaux *et al.* 2010). However, phylogenetic analyses using mitochondrial and nuclear genetic markers obtained incongruent relationships between these morphologically identified *Corbicula* species due to low bootstrap support and discordance between trees (see Hedtke *et al.* 2008, 2011; Pigneur *et al.* 2011a, 2012, 2014). The *Corbicula* taxonomy and the specific status of most freshwater lineages are therefore still debated and there is no consensus on how to name a particular *Corbicula* lineage. Even the traditional nomenclature of the four invasive androgenetic lineages, referring as form A/R, form B, form Rlc and form C/S (Siripattrawan *et al.* 2000; Lee *et al.* 2005, Marescaux *et al.* 2010) is challenged by some authors (form RlcB, Peñarrubia *et al.*, 2017; form D, Tiemann *et al.* 2017). The haploweb approach used here, delineating groups of shared alleles on a worldwide sampling dataset of *Corbicula*, including many sexual and androgenetic lineages, found only four genetic clusters (or "fields for recombination") within this genus (Fig. 2 and S3). One genetic cluster, FFR4, includes the sexual species *C. sandai* from Lake Biwa and both native and invasive androgenetic lineages (Fig. 2 and S3). These results highlight the cross-species genetic mixing and introgression events in the clam genus *Corbicula* hampering taxonomic species delimitations. Indeed, androgenesis in *Corbicula* results in hybridization and mitochondrial capture events, causing gene tree incongruences (Pigneur *et al.*, 2011a, 2014; Hedtke & Hillis, 2011; Hedtke *et al.*, 2011, Peñarrubia *et al.*, 2017, Bespalaya *et al.*, 2018) and pooling species within a same genetic cluster (Fig. 2, Fig. 3S). We therefore prefer to use the term "lineage" instead of species in *Corbicula*. While we cannot identify genetically distinct species in *Corbicula*, the allele sharing information obtained here can elucidate the geographic origin of androgenetic *Corbicula* lineages and reveal the recombination processes ongoing in androgenetic *Corbicula*.

Distinct biogeographic origins of androgenetic **Corbicula** *lineages*

Phylogenetic trees suggested separate geographic origins for the invasive lineages, with form A originating in Japan, forms B and Rlc from the Asian mainland (China, Korea, Vietnam) and form C from a *C. moltkiana*-like ancestor from Indonesia, although form C/S also clustered with other *Corbicula* lineages depending on the marker used (Siripattrawan *et al.* 2000, Park & Kim, 2003, Lee *et al.* 2005, Hedtke *et al.* 2011, Pigneur *et al.*, 2014). In this study, the haploweb approach and the conspecificity matrix gave consistent results for all markers tested.

Indeed, the haploweb approach applied on the nuclear markers found that all studied androgenetic *Corbicula* individuals cluster into three distinct genetic groups (Figures 2 & S3). Moreover, while *C. japonica* formed a distinct fourth genetic pool (confirmed by the COI network and the COI phylogenies, Fig 2d, Lee *et al.*, 2005 Glaubrecht *et al*., 2006; Hedtke *et al*. 2008, Pigneur *et al.*, 2014; Bespalaya *et al*. 2018), all the other tested sexual lineages (*C. sandai* from Japan and *C. loehensis, C. possoensis, C. matannensis and C. moltkiana* from Indonesia, Table S1) clustered with androgenetic lineages. Our analyses therefore suggest at least three distinct biogeographic origins of androgenetic lineages within the genus *Corbicula*. First, a Southeastern Asian origin for the Vietnamese androgenetic *Corbicula* lineages is suggested by the sharing of a unique α-amylase allele with one individual of the sexual Indonesian lineage *C. moltkiana* and also with the Indonesian *C. tobiae* (unknown reproductive mode) (Fig. 2, Table S1). However, the COI network and the 28S haploweb clustered the Vietnamese androgenetic individuals with the Korean androgenetic individual, while the Indonesian lineages clustered separately (Fig 2a, d). We may therefore hypothesize that there has been partial nuclear DNA capture, or a hybridization event, with the sexual *C. moltkiana* at the origin of the Vietnamese androgenetic lineages or after its origin. Since we only analyzed single individuals from these distinct locations in Indonesia and Korea, the genetic relationships with the Vietnamese individuals are still unresolved and this region needs a more thorough sampling. Second, the invasive form C/S clusters exclusively with the South African *C. fluminalis africana* individuals (whose reproductive mode could not be identified) for all four markers tested (Fig. 2, Table S1). This confirms a distinct biogeographic origin for *Corbicula* form C/S from Africa or the Middle East, the current native range of the species named *C. fluminalis* (Korniushin 2004) which was suggested by Pigneur *et al.* (2014). Third, the largest genetic pool (FFR4), includes the sexual *C. sandai* from Lake Biwa, most Asian clams and three out of the four invasive androgenetic lineages (forms A/R, B and Rlc; Fig. 2). Indeed, the alleles found in the three invasive androgenetic forms A/R, B and Rlc are found in numerous native locations, including several Asian populations (China, Korea, Taiwan, Thailand, Philippines and Japan), Madagascar, Congo and Hawaii (Table S1). Because these alleles are identical across a wide geographic distribution of *Corbicula*, the shared history appears to be recent. Within this third cluster, the population from Lake Biwa is particularly genetically diverse with numerous private alleles (Table S1) but also encompassing most alleles found in invasive lineages A/R, B, and R1c (Fig. 3). Indeed, the majority of these "invasive" alleles are found in Japan and in Lake Biwa where androgenetic lineages co-exist with the endemic sexual lineage *C. sandai* (Figures 2 and 3; Table S1). More specifically, alleles of the *amy* gene belonging to the four, distinct invasive androgenetic *Corbicula* forms, three of the four COI haplotypes (belonging to forms A/R, B and Rlc) and 28S alleles from form B and A/R were retrieved in Lake Biwa (Table S1 and Figure 3). Our analysis

therefore points to Japan, and more specifically to Lake Biwa, as a hotspot for the origin of androgenetic *Corbicula* lineages (Figure 3; Houki *et al.*, 2011). Repeated origins of asexual clones from a sexual relative have been observed in other taxa where sexual and asexual individuals co-occur (Simon *et al.*, 2003; Bode *et al.*, 2010; Neiman *et al.*, 2010). Moreover, the studies of Komaru *et al.* (2012, 2013) found that androgenetic hermaphroditic individuals of *C. leana* shared 28S rDNA alleles with sexual *C. sandai*, both co-occurring in Lake Biwa. Nevertheless, we cannot argue that *C. sandai* itself is the direct ancestor of the studied androgenetic lineages as a number of morphological, cytological and biological features (*e.g.* biflagellate *vs.* uniflagellate sperm, dioecy *vs.* hermaphroditism, non-brooding *vs.* brooding) distinguish it from androgenetic clams (Glaubrecht *et al.*, 2006). This lake however deserves a more detailed investigation.

Two patterns were hypothesized to explain the origins of androgenetic *Corbicula* lineages (Hedtke *et al.* 2011). If androgenetic lineages had a single origin from a sexual congener, followed by multiple, independent cross-lineage genetic exchanges through egg parasitism, we would expect all androgenetic clams sharing a common set of nuclear alleles and therefore constituting one unique genetic pool (or FFR) including their sexual ancestor. By contrast, androgenetic *Corbicula* lineages having multiple, independent origins from sexual congeners and evolving separately would be distinguished by distinct "fields for recombination", each containing the potential sexual ancestor (Hedtke *et al.*, 2011). Here, our results highlight three distinct genetic pools containing androgenetic lineages clustering with sexual *Corbicula* lineages or lineages whose reproductive mode still needs to be identified. While our results do not disentangle the origin of the peculiar reproductive mode of androgenesis in *Corbicula*, it shows for the first time the distinct biogeographic origins of androgenetic *Corbicula* lineages and highlights Lake Biwa, where sexual *C. sandai* occurs, as a potential hotspot for the origin of androgenetic lineages.

Genetic captures among* Corbicula *lineages

The invasive form C/S, found in Europe and South America, has an independent biogeographic origin in Africa or the Middle East. Despite belonging to a distinct allelic pool with *C. fluminalis africana*, the form C/S shares one allele (*amy* U4b) with a few individuals from Lake Biwa (Fig. 3, Table S1). The high level of divergence between this allele and the other *amy* haplotypes of form C/S (4.7%) suggests a potential genetic capture event between the Asian lineage from Lake Biwa or any androgenetic lineage carrying this allele and form C/S or its ancestor. Forms A/R and C/S also share one 28S allele (Fig. 3, Table S1) and while they appear to have distinct biogeographic origins, they are found in sympatry throughout their invasive range (*e.g.* Lee *et al.*, 2005; Hedtke *et al.*, 2008; Pigneur *et al.*, 2011, 2014). We therefore hypothesize post-introduction nuclear genetic exchanges (as

in Figure 1b) between these two invasive lineages. Individuals from invasive form B share alleles with form A/R (*amy* AA10b) and form Rlc (*amy* Db22a), the two latter also sharing an allele with each other (28S AA10a) (Fig. 3 and Table S1). Hedtke *et al.* (2008) already observed genome captures between forms A and B. The 'form D' described in Tiemann *et al.* 2017 may also be a hybrid between these two forms. Furthermore, a sampling of invasive *Corbicula* lineages in the Iberian Peninsula revealed a large amount of heterozygous individuals displaying 28S alleles from form A/R associated to one allele of form Rlc or form C/S, sometimes even associated to the COI haplotype of the third lineage (Peñarrubia *et al.* 2017). Recently, in a man-made channel in Northern European Russia, all sampled *Corbicula* individuals belonged to two distinct invasive lineages apparently fixed for a mismatch between the mtDNA haplotype and morphotype. The first lineage belonged to morphotype Rlc associated to the FW5 haplotype of form A/R, the second lineage had morphotype A/R associated to FW17 COI haplotype of form C/S (Bespalaya *et al*. 2018). Out of twelve sampled locations in Europe where form A/R occurs in sympatry with another invasive form (C/S or Rlc), cytonuclear mismatches were detected at 8 locations and hybrids at 2 locations (Pigneur *et al.* 2014; Etoundi, pers. obs.).

While the genetic diversity within the four androgenetic *Corbicula* forms in the invaded range in Europe and America is very low (Pigneur *et al.*, 2014), androgenetic *Corbicula* lineages found in the native range, particularly Asia, are both genetically and morphologically diverse (*e.g.* Glaubrecht *et al.*, 2003; Hedtke *et al.*, 2011; Pigneur *et al.*, 2014; Table S1). Interestingly, the most widespread alleles (28S AA10a and *amy* AA10b, Table S1) shared between all lineages and populations from FFR4 (Figures 2 and 3), belong to the invasive form A/R, being the widespread clone both in America and Europe (Pigneur *et al.* 2014). Moreover, *Corbicula* from Madagascar and Hawaii also shared the same allele. Two hypotheses could explain this pattern: (i) these widespread alleles represent shared ancestral alleles already present before the invasion and divergence towards other Asian countries, America and Europe; (ii) these alleles are signatures of nuclear genetic exchanges in contact zones. In addition, the Hawaiian population shared alleles with several Chinese populations (BA, Muj and FU) suggesting an invasion from China to Hawaii (Fig. 3, Table S1).

We also found that both the invasive and native androgenetic *Corbicula* individuals are mostly heterozygotes (also confirmed by Pigneur *et al.*, 2014; Peñarrubia et al. 2017). Hybridization or nuclear genetic capture events between lineages, which are more likely to occur in asexual organisms that retained sexual features (such as androgens where fertilization occurs, Figure 1), might facilitate the evolution of genetic diversity, enhanced heterozygosity and the occurrence of distinct alleles from different genetic lineages within one individual. Hybridization could be facilitated in an androgenetic system where distinct genetic lineages co-occur, but also in

populations where sexuals and asexuals are found in sympatry (such as Lake Biwa in Japan). It has indeed been demonstrated in *Corbicula* that egg parasitism of a distinct androgenetic lineage could be accompanied not only by capture of maternal mitochondrial DNA, but also of nuclear DNA, increasing the ploidy and the diversity of hybrids (Figure 1) (Hedtke *et al*. 2011, Peñarrubia *et al*. 2017, Bespalaya *et al*. 2018). The ancestral polymorphism (and heterozygosity) hereby obtained, is retained in asexuals and results in divergent alleles within individuals. Since many androgenetic *Corbicula* individuals are polyploid, we may hypothesize that many androgenetic *Corbicula* lineages appear to have originated through hybridization, or more particularly through partial nuclear captures following egg parasitism.

Origin of androgenesis in *Corbicula* yet unresolved

While hybridization has been demonstrated between distinct androgenetic *Corbicula* lineages (*e.g.* Pigneur *et al.* 2014), it is not known yet whether androgenesis in *Corbicula* has a single origin by one or more mutations affecting meiosis and spermatogenesis, or multiple origins as a result of repeated mutations or hybridization events between distinct sexual lineages (Hedtke *et al.* 2011; Pigneur *et al.*, 2012, Lehtonen *et al.*, 2013). Hedtke & Hillis (2011) suggest that mutations in genes for spindle fiber formation during meiosis could cause the evolution of androgenetic reproduction within the genus *Corbicula*, as this could both lead to unreduced sperm and the formation of polar bodies containing the entire maternal genome after fertilization (Figure 1). Even with a single origin, the ancestral androgenetic clone could spread and capture genomic portions of other *Corbicula* lineages, radiating and forming distinct androgenetic lineages. Contemporary populations of *Corbicula* are dominated by androgenetic lineages, and there are few populations where distinct sexual lineages live in sympatry. Again, Lake Biwa is of peculiar interest to investigate the transition mechanism in *Corbicula* between sexuality and androgenesis and our study here calls for further investigations of this lake system, while it is not the only region where androgenetic lineages appear to originate (as shown in this study).

Acknowledgments

We thank B. Hespeels, M. Terwagne, L. Bankers, K. McElroy, J. Sharbrough, M. Neiman, L. Keller and T. Schwander for advice on the analyses and proofreading, as well as N. Yasuda and the late M. Yamaguchi for access to Japanese samples, particularly specimens from Lake Biwa. We are also thankful to all the colleagues who provided samples (C. Ayres, Y. Bagatini, L. Beran, D. Eekelers, J. Higuti, C. Ituarte, K. de Koch, P. Laforge, T. Lee, S. Milla, K. Martens, J. Mouthon, N. Perrin, M.-A. Pierrard, S. Schmidlin, R. Sousa, N. Spann, E. Tinti, M. Yamaguchi and the Illinois Natural History Survey). This work was supported by the Belgian National Fund for Scientific Research (Fonds National pour la Recherche Scientifique - FNRS) through a PhD thesis grant (FC91712) to E.E. and by a PDR research grant (CARMA, 14596412) to K.V.D. and by UNamur for the teaching assistant grant to M.V.

Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**:716-723

Araujo, R., Moreno, D., Ramos, M.A. 1993. The Asiatic clam *Corbicula fluminea* (Müller, 1974) (Bivalvia: Corbiculidae) in Europe. *American Malacological Bulletin*. **10**(1):39-49

Bah, T. 2011. Inkscape. Guide to a vector drawing program. 4th edition. Prentice Hall

Bandelt HJ, Forster P. & Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37-48

Barbuti, R., Mautner, S., Carnevale, G., Milazzo, P., Rama, A. & Sturmbauer, C. 2012. Population dynamics with a mixed type of sexual and asexual reproduction in a fluctuating environment. *BMC Evolutionary Biology* **12**:49

Bespalaya, Y.V., Bolotov, I.N., Aksenova, O.V., Kondakov, A.V., Gofarov, M.Y., Laenko, T.M., Sokolova, S.E., Shevchenko, A.R., Travina, O.V. 2018 Aliens are moving to the Arctic frontiers: an integrative approach reveals selective expansion of androgenic hybrid *Corbicula* lineages towards the North of Russia. *Biol. Invasions*. https://doi.org/10.1007/s10530-018-1698-z

Bode, S.N.S., Adolfsson, S., Lamatsch, D.K., Martins, M.J.F., Schmit, O., Vandekerkhove, J. et al. 2010. Exceptional cryptic diversity and multiple origins of parthenogenesis in a freshwater ostracod. *Molecular Phylogenetics and Evolution* **54**:542-552

Britton, J.C. & Morton, B. 1986. Polymorphism in *Corbicula fluminea* (Bivalvia: Corbiculoidea) from North *American Malacological Review* **19**:1–43

Castagnone-Sereno, P. 2006. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity* **96**:282–289

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772

Debortoli, N., Li, X., Eyres, I., Fontaneto, D., Hespeels, B., Tang, C. Q., Flot, J.-F., & Van Doninck, K. 2016. Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. *Current Biology* **26**(6), 723-732.

Delmotte, F., Sabater-Muñoz, B., Prunier-Leterme, N., Latorre, A., Sunnucks, P., Rispe, C. & Simon, J.-C. 2003. Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution* **57**:1291-1303

Doyle, J.J. 1995. The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. *Systematic Botany* **20**:574-588

Felsenstein, J. 1974. The evolutionary advantage of recombination. *Genetics* **1**:737-756

Fisher, R.A. 1930. The genetical theory of natural selection. Clarendon, Oxdord

Flot, J.F., Tillier, A., Samadi, S. & Tillier, S. 2006. Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Resources* **6**:627-630

Flot, J.F. 2007. Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Resources* **7**:974-977

Flot, J.F. 2010. SeqPHASE: a web tool for interconverting PHASE input⁄output files and FASTA sequence alignments. *Molecular Ecology Resources* **10**:162–166

Flot, J.F., Couloux, A. & Tillier, S. 2010. Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. *BMC Evolutionary Biology* **10**:372

Fontaneto, D., Flot J.-F. & Tang, C.Q. 2015. Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity* **45**:433-451

Glaubrecht, M., von Rintelen, T. & Korniushin, A.V. 2003. Toward a systematic revision of brooding freshwater Corbiculidae in Southeast Asia (Bilvavia, Veneroida): on shell morphology, anatomy and molecular phylogenetics of endemic taxa from islands in Indonesia. *Malacologia* **45**:1-10

Glaubrecht, M., Fehér, Z. & von Rintelen, T. 2006. Brooding in *Corbicula madagascariensis* (Bivalvia, Corbiculidae) and the repeated evolution of viviparity in corbiculids. *Zoologica Scripta* **35**:641-654

Gomes C., Sousa R., Mendes T., Borges R., Vilares P., Vasconcelos V., Guilhermino, L., Antunes, A. 2016 Low genetic diversity and high invasion success of *Corbicula fluminea* (Bivalvia, Corbiculidae) (Müller, 1774) in Portugal. *PLoS ONE* **11**(7): e0158108. doi:10.1371/journal.pone.0158108

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Research* **41**:95-98

Hedtke, S.M., Stanger-Hall, K., Baker, R.J. & Hillis, D.M. 2008 All-male asexuality: origin and maintenance of androgenesis in the Asian clam *Corbicula*. *Evolution* **62**:1119-1136

Hedtke, S.M. 2009. Origin and maintenance of androgenesis: male asexual reproduction in the clam genus *Corbicula*. Thesis dissertation, University of Texas at Austin (129 pages)

Hedtke, S.M., Glaubrecht, M. & Hillis, D. 2011. Rare gene capture in predominantly androgenetic species. *Proc. Natl. Acad. Sci. U.S.A.* **108**:9520-8524

Hedtke, S.M. & Hillis, D. 2011. The potential role of androgenesis in cytoplasmic–nuclear phylogenetic discordance. *Systematics Biology* **60**:87-109

Hillis, D.M. & Patton, J.C. 1982. Morphological and electrophoretic evidence for two species of *Corbicula* (Bivalvia: Corbiculidae) in North America. *American Midland Naturalist* **108**:74-80

Hotta M. & Komaru A. 2018. The process of first polar body formation in eggs of the androgenetic clam *Corbicula fluminea*. *Journal of Shellfish Research* **37:** 131-137.

Houki, S., Yamada, M., Honda, T. & Komaru, A. 2011. Origin and possible role of males in hermaphroditic androgenetic *Corbicula* clams. *Zoological Science* **28**:526-531

Huelsenbeck, J.P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754-755

Hurvich, C.M. & Tsai, C.L. 1989. Regression and time series model selection in small samples. *Biometrika* **76**:297-307.

Ishibashi, R., A. Komaru, K. Ookubo & M. Kiyomoto. 2002. The second meiosis occurs in cytochalasin D-treated eggs of Corbicula leana even though it is not observed in control androgenetic eggs because the maternal chromosomes and centrosomes are extruded at first meiosis. *Dev. Biol.* **244:** 37–43

Ishibashi, R., Ookubo, K., Aoki, M., Utaki, M., Komaru, A. & Kawamura, K. 2003. Androgenetic reproduction in a freshwater diploid clam *Corbicula fluminea* (Bivalvia: Corbiculidae). *Zool. Sci.* **20**:727-732

Katoh, K., Rozewicki, J., Yamada, K.D. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, bbx108, https://doi.org/10.1093/bib/bbx108

Komaru, A., Konishi, K., Nakayama I., Kobayashi, T., Sakai, H., & Kawamura, K. 1997. Hermaphroditic freshwater clams in the genus *Corbicula* produce non-reductional spermatozoa with somatic DNA content. *Biology Bulletin* **193**:320-323

Komaru, A. & Konishi, K. 1999. Non-reductional spermatozoa in three shell color types of the freshwater clam *Corbicula fluminea* in Taiwan. *Zoological Science* **16**:105-108

Komaru, A., Ookubo, K. & Kiyomoto, M. 2000. All meiotic chromosomes and both centrosomes at spindle pole in the zygotes discarded as two polar bodies in clam *Corbicula leana*: unusual polar body formation observed by antitubulin immunofluorescence. *Developmental Genes and Evolution* **210**:263-269

Komaru, A., Kumamoto, A., Kato, T., Ishibashi, R., Obata, M. & Nemoto, T. 2006. A hypothesis of ploidy elevation by formation of a female pronucleus in the androgenetic clam *Corbicula fluminea* in the Tone River estuary, Japan. *Zoological Science* **23**: 529-532

Komaru, A., Houki, S., Yamada, M., Miyake, T., Obata, M., Kawamura, K. 2012. 28S rDNA haplotypes of males are distinct from those of androgenetic hermaphrodites in the clam *Corbicula leana*. *Dev. Genes Evol*. **222**:181–187

Komaru, A., Yamada, M., Houki, S. 2013. Relationship Between Two Androgenetic Clam Species, *Corbicula leana* and *Corbicula fluminea*, Inferred from Mitochondrial Cytochrome b and Nuclear 28S rRNA Markers. *Zoological Science*. **30**(5):360-365.

Kondrashov, A.S. 1988. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**:435-440

Korniushin, A.V. 2004. A revision of some Asian and African freshwater clams assigned to *Corbicula fluminalis* (Müller, 1774) (Mollusca: Bivalvia: Corbiculidae), with a review of anatomical characters and reproductive features based on museum collections. *Hydrobiologia* **529**:251-270

Kraemer, L.R., Swanson, C., Galloway, M. & Kraemer, R. 1986. Biological basis of behavior in *Corbicula fluminea*. II. Functional morphology of reproduction and development and review of evidence for self-fertilization. *American Malacolical Bulletin* Spec. Ed., **2**:193-202.

Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne R, Horsman, D., et al. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* **19**:1639-1645

Lee, T., Siripattrawan, S., Ituarte, C.F. & Ò Foighil, D. 2005. Invasion of the clonal clams: *Corbicula* lineages in the New World. *American Malacological Bulletin.* **20**:113-122

Lehtonen, J., Schmidt, D.J., Heubel, K. & Kokko, H. 2013. Evolutionary and ecological implications of sexual parasitism. *Trends in Ecology and Evolution* **28**: 297-306

Librado, P. & Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451-1452

Mantovani, B., Passamonti, M. & Scali, V. 2001. The mitochondrial cytochrome oxidase II gene in *Bacillus* stick insects: ancestry of hybrids, androgenesis, and phylogenetic relationships. *Molecular Phylogenetics and Evolution* **19**:157–163.

Marescaux, J., Pigneur, L.M. & Van Doninck, K. 2010. New records of Asian clams *Corbicula* spp. in French Rivers. *Aquatic Invasions* **5**:S35-S39

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**(1): vev003

Maynard Smith, J. 1978. *The Evolution of Sex*. Cambridge University Press, Cambridge.

McMahon, R.F. 1982. The occurrence and spread of the introduced Asiatic freshwater clam, *Corbicula fluminea* (Muller), in North America: 1924–1982. *Nautilus*. **96**:134–141

Mouthon, J. 1981. Sur la présence en France et au Portugal de *Corbicula* (Bivalvia, Corbiculidae) originaire d'Asie. *Basteria*. **45**:109–116

Muller, H.J. 1932. Some genetic aspects of sex. *American Naturalist* **66**:118-138

Muller, H.J. 1964. The relation of recombination to mutational advance. *Mutation Research.* **1**:2–9

Neiman, M., Hehman, G., Miller, J.T., Logsdon J.M., Jr. & Taylor, D.R. 2010. Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Molecular Biology and Evol*ution**27**:954–963

Neiman, M., Schwander, T. 2011. Using parthenogenetic lineages to identify advantages of sex. *Evolutionary Biol*ogy **38**: 115-123

Otto, S.P., 2008. Sexual reproduction and the evolution of sex. *Nature Education* **1**:182

Paczesniak, D., Jokela, J., Larkin, K. & Neiman, M. 2013. Discordance between nuclear and mitochondrial genomes in sexual and asexual lineages of the freshwater snail *Potamopyrgus antipodarum*. *Molecular Ecology* **22**: 4695-4710

Park, J.K. & Kim, W. 2003. Two *Corbicula* (Corbiculidae: Bivalvia) mitochondrial lineages are widely distributed in Asian freshwater environment. *Molecular Phylogenetics and Evolution* **29**: 529-539

Peñarrubia, L., Araguas, R.M., Vidal, O., Pla, C., Viñas, J., Sanz, N. 2017. Genetic characterization of the Asian clam species complex (*Corbicula*) invasion in the Iberian Peninsula. *Hydrobiologia*. DOI 10.1007/s10750-016-2888-2

Pfenninger, M., Reinhardt, F. & Streit, B. 2002. Evidence for cryptic hybridization between different evolutionary lineages of the invasive clam genus *Corbicula* (Veneroida, Bivalvia). *Journal of Evolutionary Biology* **15**: 818-829

Pigneur, L.-M., Marescaux, J., Roland, K., Etoundi, E., Descy J.-P. & Van Doninck, K. 2011a. Phylogeny and androgenesis in the invasive *Corbicula* clams (Bivalvia, Corbiculidae) in Western-Europe. *BMC Evolutionary Biology.* **11**:147.

Pigneur, L.-M., Risterucci, A.-M., Dauchot, N., Li, X., Van Doninck, K. 2011b. Development of novel microsatellite markers to identify the different invasive lineages in the *Corbicula* complex and to assess androgenesis. *Molecular Ecology Resources*, **11**: 573–577.

Pigneur, L.-M., Hedtke, S.M., Etoundi, E. & Van Doninck, K. 2012. Androgenesis: a review through the study of the selfish shellfish *Corbicula* spp. *Heredity* **108**:581-591

Pigneur, L.-M., Etoundi, E., Aldridge, D.C., Marescaux, J., Yasuda, N. & Van Doninck, K. 2014. Genetic uniformity and long-distance clonal dispersal in the invasive androgenetic *Corbicula* clams. *Molecular Ecology* **23**:5102-16

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Renard, E., Bachmann, V., Cariou, M.L., Moreteau, J.C. 2000. Morphological and molecular differentiation of invasive freshwater species of the genus *Corbicula* (Bivalvia, Corbiculidea) suggest the presence of three taxa in French rivers. *Molecular Ecology*. **9**:2009–2016.

Rintelen, T. v. & Glaubrecht, M. 2006. Rapid evolution of sessility in an endemic species flock of the freshwater bivalve *Corbicula* from ancient lakes on Sulawesi, Indonesia. *Biology Letters*. **2**:73–77

Ronquist, F., Huelsenbeck, J. & Teslenko, M. 2001. Draft MrBayes version 3.2 Manual: tutorials and model summaries.

Simon, J.C., Delmotte, F., Rispe, C. & Crease, T. 2003. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society* **1**: 151-163

Schwander, T., & Crespi B.J. 2009. Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in *Timema* stick insects. *Evolution* **63**:84-103

Schwander, T., Henry, L. & Crespi B.J. 2011. Molecular evidence for ancient asexuality in *Timema* stick insects. *Current Biology* **21**: 1129-1134

Siripattrawan, S., Park, J.-K. & Ò Foighil, D. 2000. Two lineages of the introduced Asian freshwater clam *Corbicula* occur in North America. *Journal of Molluscan Studies* **66**: 423–429

Stephens, M., Smith, N. & Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *American* Journal of *Human Genetics* **68**: 978-989

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**: 2731-2739

Tiemann, J.S., Haponski, A.E., Douglass, D.A., Lee, T., Cummings, K.S., Davis, M.A., Foighil, D.Ó. 2017. First record of a putative novel invasive Corbicula lineage discovered in the Illinois River, Illinois, USA. *BioInvasions Records*. **6**(2):159-166

Yamada, M., Ishibashi, R., Toyoda, K., Kawamura, K. & Komaru, A. 2014. Phylogeography of the brackish water clam *Corbicula japonica* around the Japanese archipelago inferred from mitochondrial COII gene sequences. *Zoological Science* **31**: 168-179

Zhao, S., Guo, Y., Sheng, Q., & Shyr, Y. 2014 Advanced Heat Map and Clustering Analysis Using Heatmap3. *BioMed Research International*. Article ID 986048, 6p. https://doi.org/10.1155/2014/986048.

Data Accessibility

- Sample information: Table S1
- DNA sequences: Genbank accessions (Table S2)

Author contribution

E.E. and K.V.D. designed the questions and research. E.E., and J.V. performed the research. E.E., J.M., N.D., M.V. and K.V.D. analyzed the data. J.F.F. contributed to new analytical tool development. L.-M.P. and S.M.H. provided sequences and samples. E.E., N.D., J.M., M.V. and K.V.D. made the figures and tables. E.E., J.M., L.-M.P., M.V. and K.V.D. wrote the paper. K.V.D. supervised the work and provided funding.

Figure captions

Figure 1: Schematic representation of the outcome of androgenetic reproduction between two distinct genetic *Corbicula* lineages (blue and red) (adapted from Pigneur *et al.* 2012). a) Offspring inherits the nuclear genome of the paternal lineage (red) and the mitochondrial genome of the maternal genome (blue), leading to a cytonuclear mismatch while the maternal nuclear chromosomes are expulsed as two polar bodies following fertilization (see Komaru *et al*. 1998). b) When maternal nuclear chromosomes (blue) are partially retained, the offspring will have a mixed (and usually polyploid) nuclear genome inherited from both maternal and paternal lineages (Komaru *et al*. 2006; Hedtke *et al*. 2011). The gamete size is not to scale.

Figure 2: 28S (a), *amy* (b) and *atps* (c) haplowebs and COI (d) haplonet. The size of each circle is proportional to the frequency of the represented alleles/haplotypes (singletons were omitted). The number of mutational steps inferred by the median-joining algorithm is displayed on the lines connecting the haplotypes. For haplowebs, the curves connecting each pair of alleles co-occurring in heterozygous individuals are drawn (width proportional to the number of individuals in which the two alleles co-occur). The distinct fields of recombination (FFR) are highlighted in different colors. Name and origin of *Corbicula* lineages follow Table S1. The few individuals clustering individually are shown in light grey.

Figure 3: Allele sharing among *Corbicula*. Groups are represented by circle arcs with a size determined by the number of alleles found in each group. Linking lines between these groups represent shared alleles, with these from *C.* sp. form A/R, B, C/S and Rlc respectively in red, blue, yellow and green.

Table 1: Genetic diversity indices: haplotype diversity (Hd) and nucleotide diversity (π) for all *Corbicula* samples based on COI, 28S, *amy* and *atps*.

Supporting information

Supporting information 1: PCR conditions for all markers used in this study.

Figure S1: Schematic description of the haploweb method.

Figure S2: 28S haploweb without hybrid individuals.

Figure S3: Conspecificity matrix. Individuals are plotted against each other and conspecificity scores are attributed to each pair of individuals. These scores correspond to the number of times individuals were considered belonging to the same FFR, based on haploweb approach, and thus range from 0 (individuals were never found on the same FFR) to 4 (individuals belonging to the same FFR for all markers). Individuals are sorted using hierarchical clustering.

Table S1: List of the 447 individuals collected for the present study and genetic results obtained for the 348 analyzed individuals. The color code corresponds to the four invasive lineages: alleles retrieved from *C.* sp. form A/R are in red, these from *C.* sp. form B are in blue, these from *C.* sp. form C/S are in yellow and these from *C.* sp. form Rlc are in green. Shared allele from one population to another are in bold.

Table S2: GenBank accession numbers, haplotype designation, taxon, reproductive mode and localities of the *Corbicula* spp. sequences used in the present study.

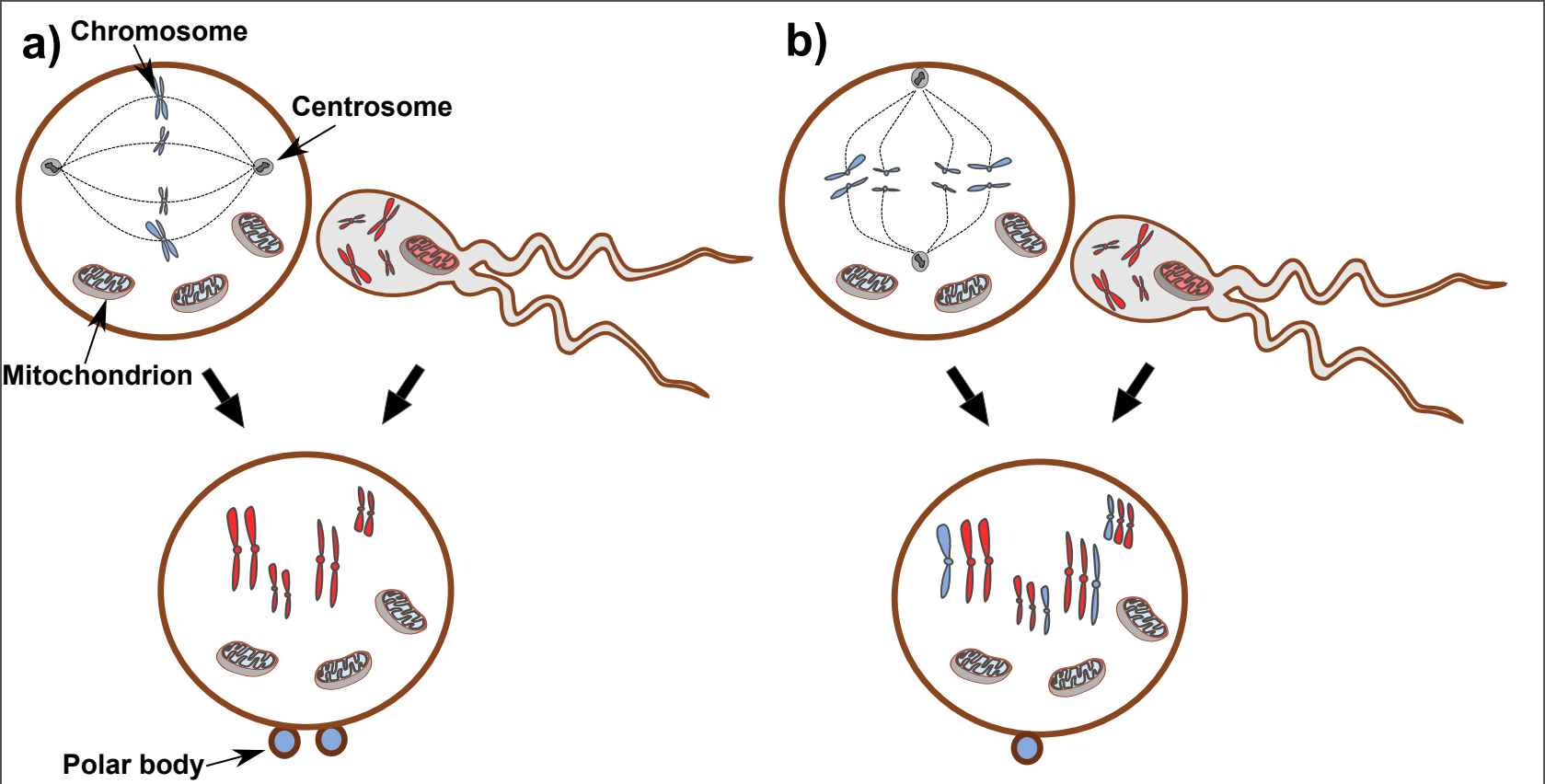

Figure 1: Schematic representation of the outcome of androgenetic reproduction between two distinct genetic *Corbicula* lineages (blue and red). a) Cytonuclear mismatch. b) Hybridization. The gamete size is not to scale.

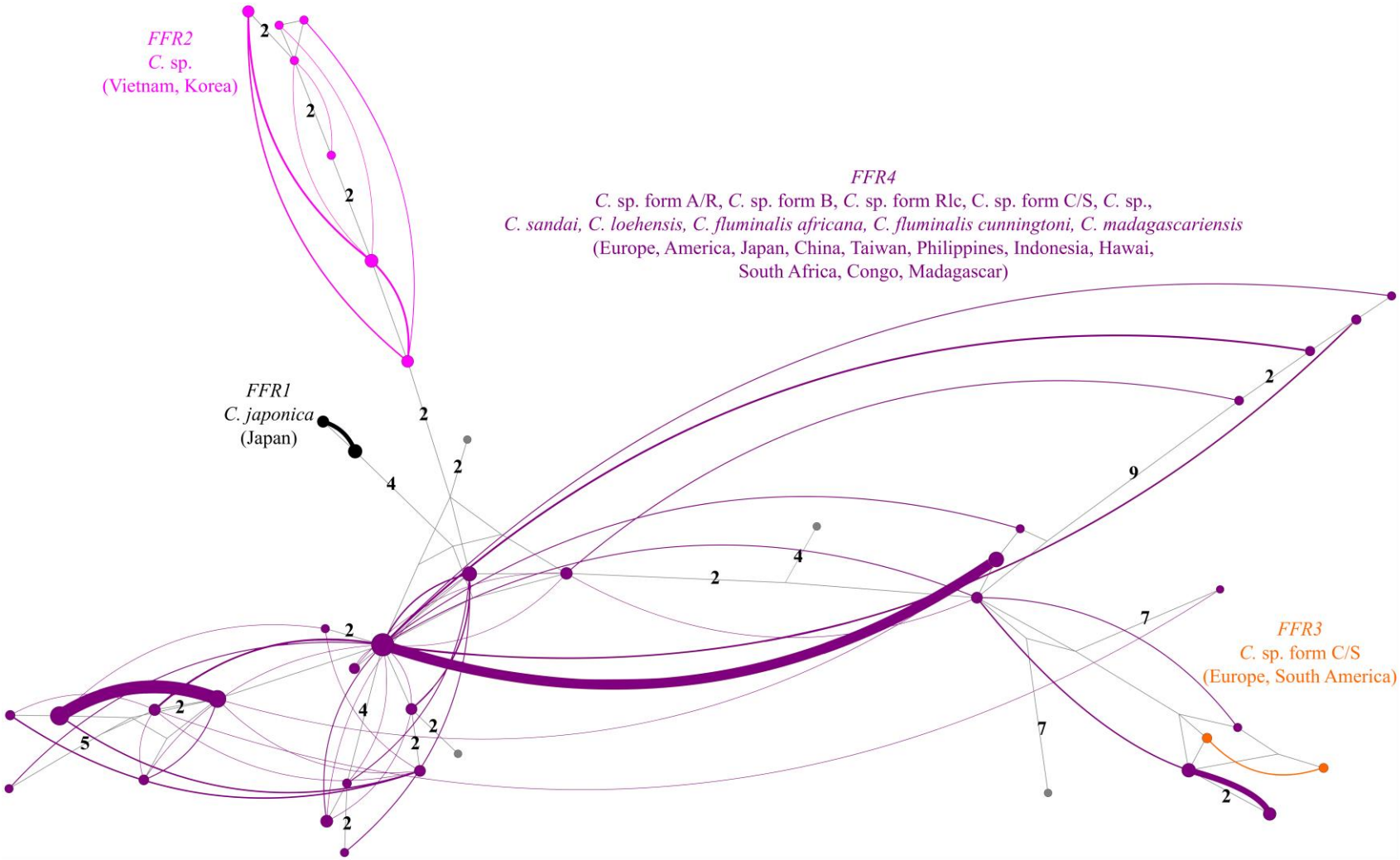

Figure 2a: 28S haploweb. Median-joining haplotype network (haplonet) based on sequence alignments, with circle size proportional to the frequency of the represented alleles/haplotypes (singletons omitted). The number of mutation steps inferred by the median-joining algorithm is displayed on the lines connecting the haplotypes. The haplonet was subsequently converted into a haplotype web (haploweb) by drawing curves connecting each pair of alleles found co-occurring in heterozygous individuals, with a width proportional to the number of individuals in which the two alleles co-occur. This allowed the delineation of fields for recombination (FFRs), *i.e.* groups of individuals sharing a unique pool of alleles. These allele pools are highlighted in different colors. The few individuals clustering individually are shown in light grey. The name and origin of each *Corbicula* lineage follows the nomenclature of Table S1.

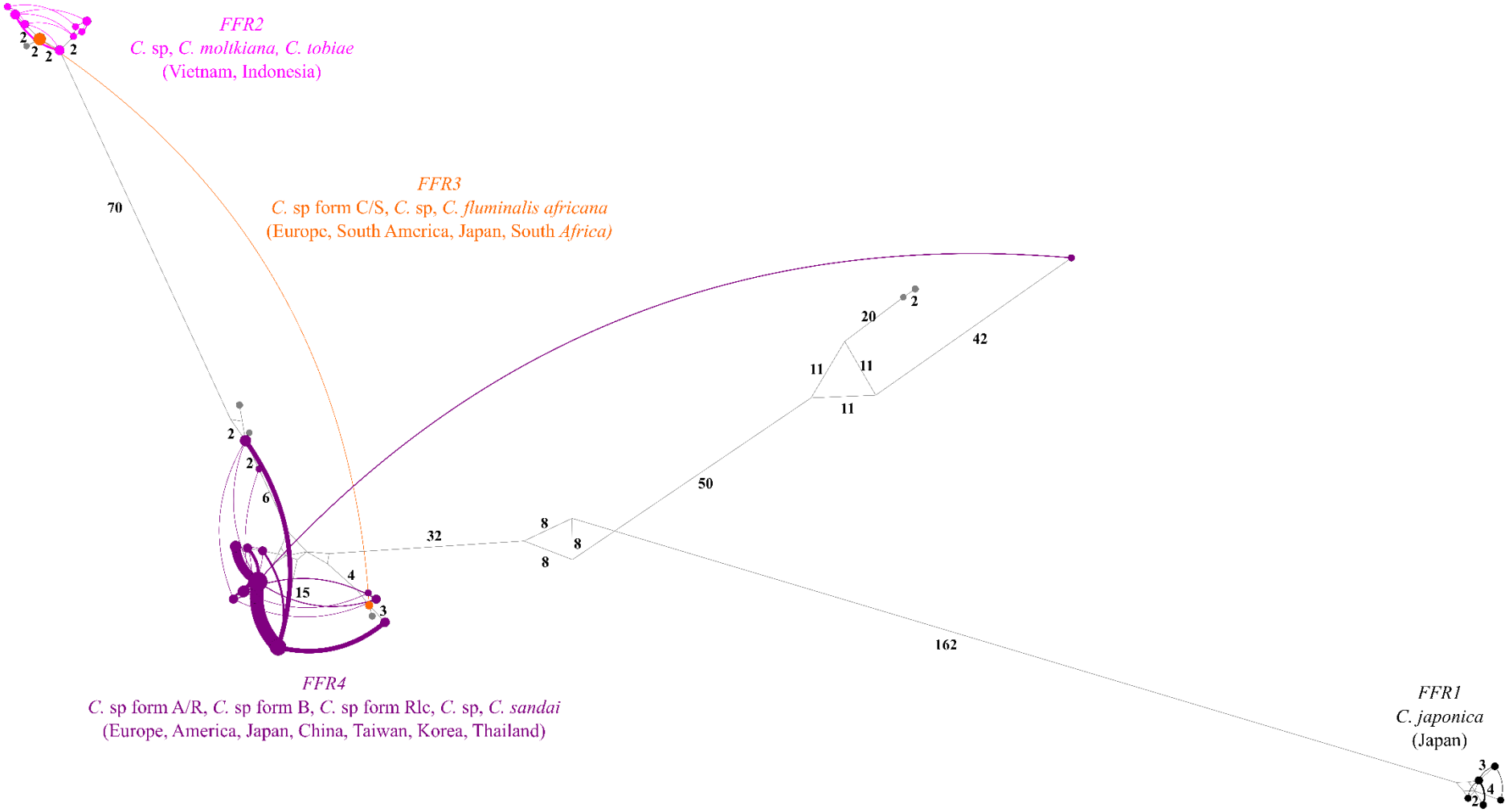

Figure 2b: *amy* haploweb. Median-joining haplotype network (haplonet) based on sequence alignments, with circle size proportional to the frequency of the represented alleles/haplotypes (singletons omitted). The number of mutation steps inferred by the median-joining algorithm is displayed on the lines connecting the haplotypes. The haplonet was subsequently converted into a haplotype web (haploweb) by drawing curves connecting each pair of alleles found co-occurring in heterozygous individuals, with a width proportional to the number of individuals in which the two alleles co-occur. This allowed the delineation of fields for recombination (FFRs), *i.e.* groups of individuals sharing a unique pool of alleles. These allele pools are highlighted in different colors. The few individuals clustering individually are shown in light grey. The name and origin of each *Corbicula* lineage follows the nomenclature of Table S1.

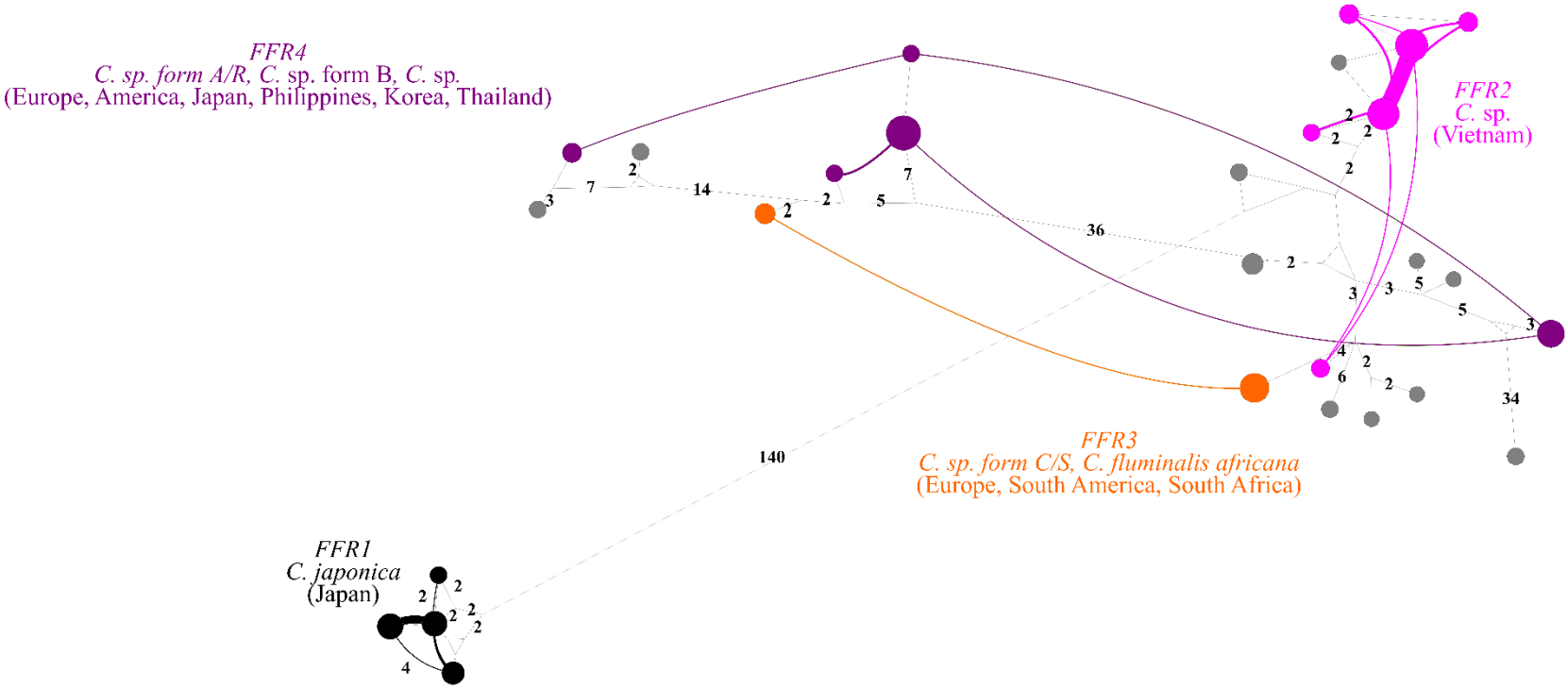

Figure 2c: *atps* haploweb. Median-joining haplotype network (haplonet) based on sequence alignments, with circle size proportional to the frequency of the represented alleles/haplotypes (singletons omitted). The number of mutation steps inferred by the median-joining algorithm is displayed on the lines connecting the haplotypes. The haplonet was subsequently converted into a haplotype web (haploweb) by drawing curves connecting each pair of alleles found co-occurring in heterozygous individuals, with a width proportional to the number of individuals in which the two alleles co-occur. This allowed the delineation of fields for recombination (FFRs), *i.e.* groups of individuals sharing a unique pool of alleles. These allele pools are highlighted in different colors. The few individuals clustering individually are shown in light grey. The name and origin of each *Corbicula* lineage follows the nomenclature of Table S1.

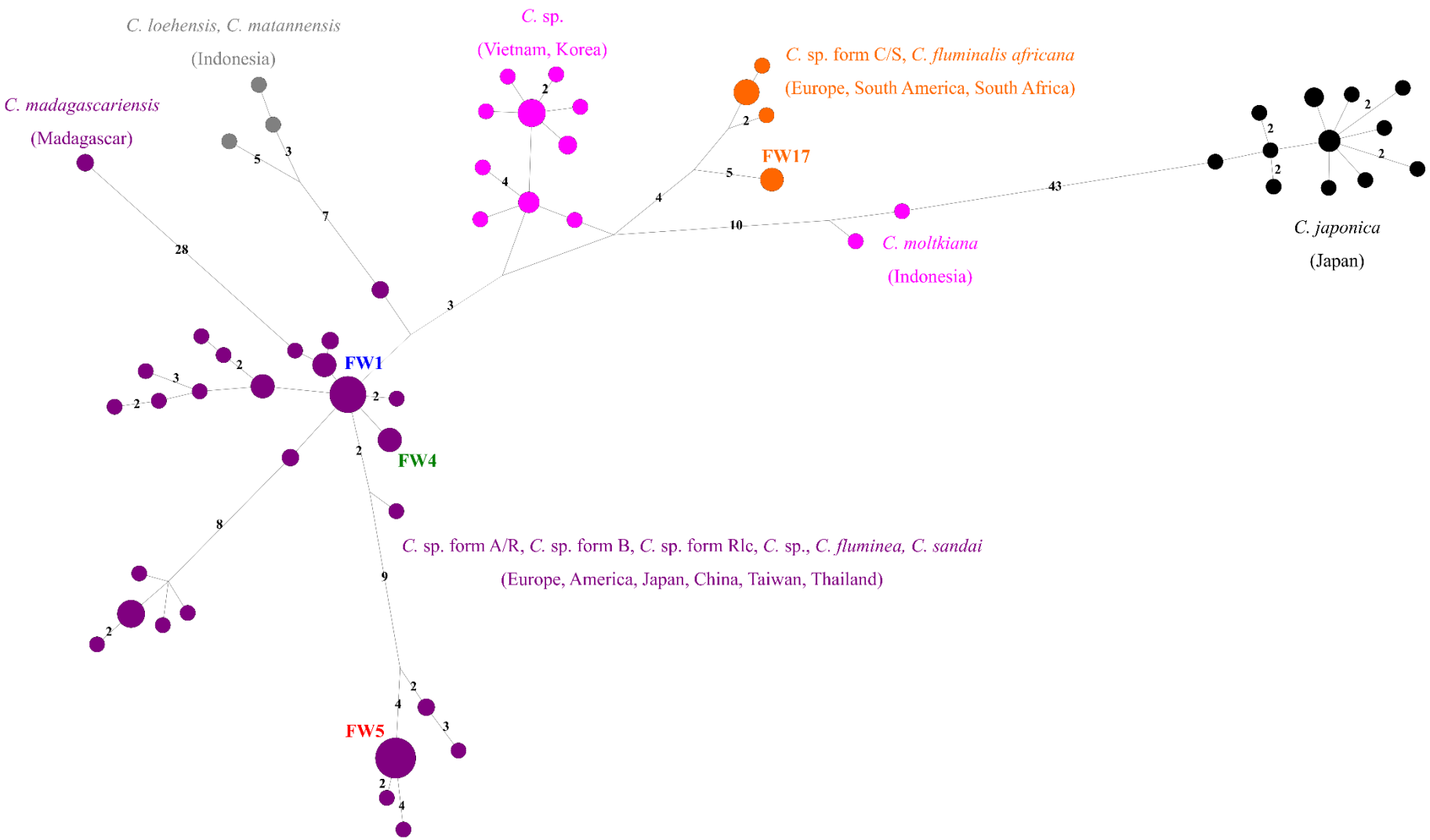

Figure 2d: COI haplonet. Median-joining haplotype network (haplonet) based on sequence alignments, with circle size proportional to the frequency of the represented alleles/haplotypes. The number of mutation steps inferred by the median-joining algorithm is displayed on the lines connecting the haplotypes. The colors used correspond to the FFR delimitation from the previous haplowebs, based on nuclear gene markers. The few individuals clustering individually are shown in light grey. The name and origin of each *Corbicula* lineage follows the nomenclature of Table S1.

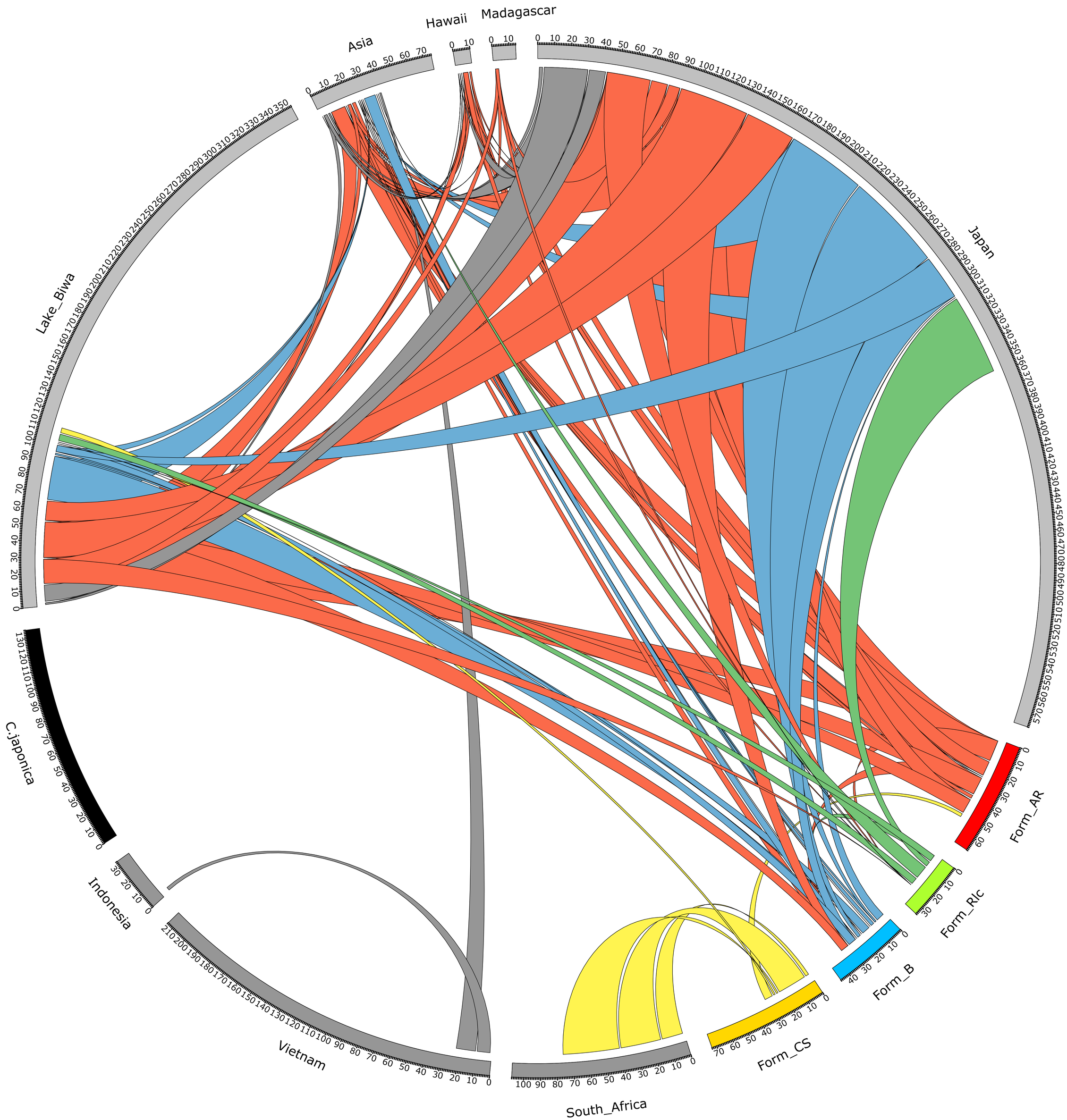

Figure 3: Circos plot representing allele sharing among *Corbicula*. Groups are represented by circle arcs with a size determined by the number of alleles found in each group. Linking lines between these groups represent shared alleles, with these from *C.* sp. form A/R, B, C/S and Rlc in red, blue, yellow and green respectively.

Table 1

Haplotype and nucleotide diversity of *Corbicula* spp. based on COI, 28S, *amy* and *atps* markers

	COI	**28S**	***amy***	***atps***
Haplotype diversity (Hd)	0.909 ± 0.010	0.953 ± 0.004	0.922 ± 0.007	0.956 ± 0.004
Nucleotide diversity (π)	0.030 ± 0.002	0.018 ± 0.001	0.050 ± 0.004	0.148 ± 0.008