

Supplementary material for:

**Genetic sex determination in three closely related hydrothermal vent gastropods, one of which has intersex individuals**

Castel J, Pradillon F, Cueff V, Leger G, Daguin-Thiébaud C, Ruault S, Mary J, Hourdez S, Jollivet D, and Broquet T

## Sampling

Table S1 -Number of females, males, and unsexed individuals per sampling location. In *Alviniconcha kojimai*, column F or H indicates the number of females (no male tissue visible in the histological section) and morphological hermaphrodites with varying proportions of male tissue. The sex indicated in this table is the sex observed from histological analyses (most reliable method) for 170 individuals, and the sex observed from the external colour of the gonad for the remaining 106 individuals.

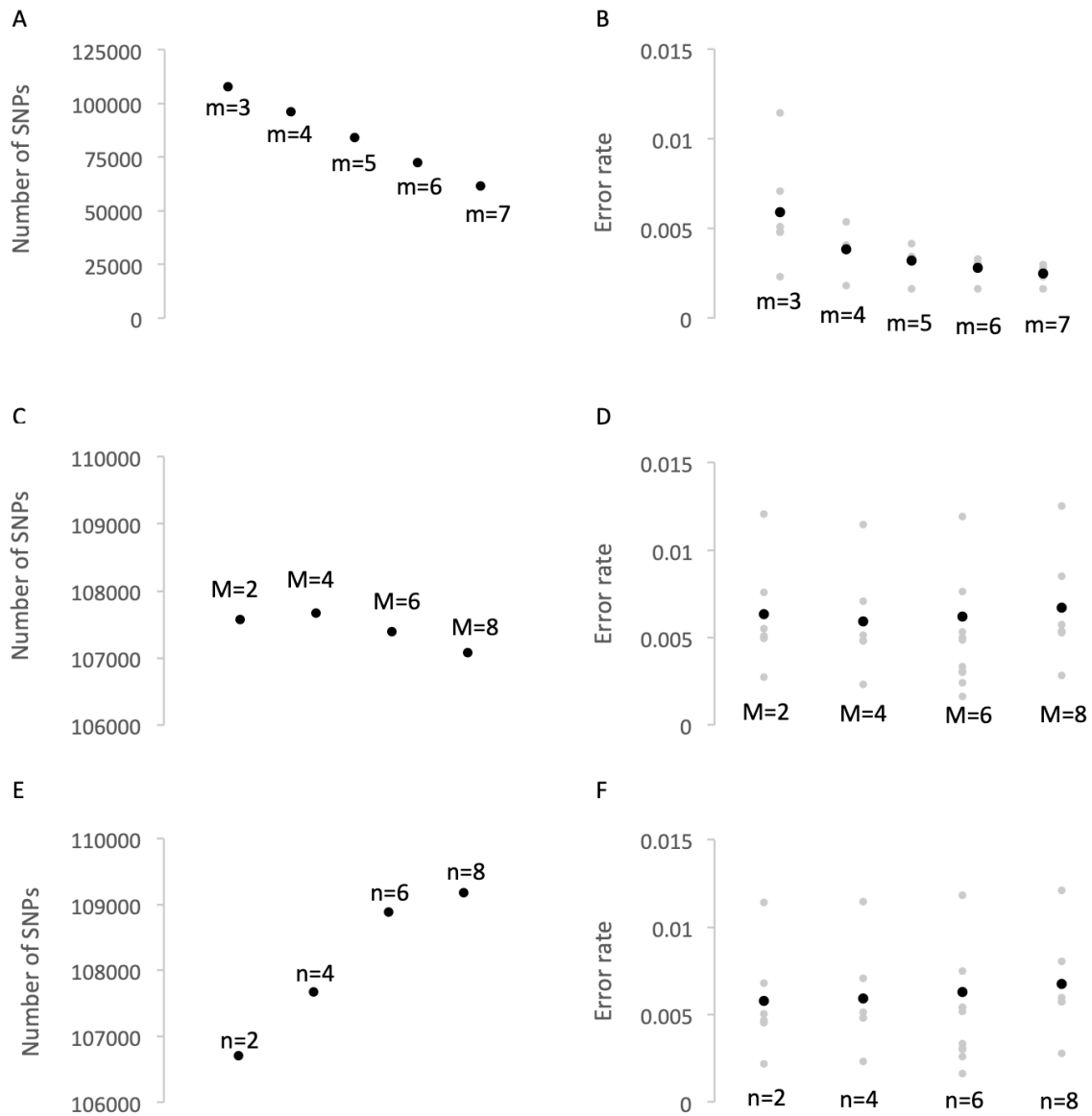
Basin or volcanic arc	Vent field	<i>A. boucheti</i>			<i>A. strummeri</i>			<i>A. kojimai</i>		
		F	M	U	F	M	U	F or H	M	U
Manus	Pacmanus	18	13	46	-	-	-	-	-	-
	Susu	1	2	47	-	-	-	6	10	31
Woodlark	La Scala	9	10	1	-	-	-	2	3	19
North Fiji	Phoenix	-	-	-	3	3	-	11	14	6
Futuna	Fati-Ufu	2	2	-	1	2	1	16	13	7
	Fatu-Kapa	-	-	-	-	-	-	6	11	-
Lau	Mangatolo	9	9	-	-	-	-	9	9	-
	Tow Cam	3	6	9	-	-	-	4	7	23
	ABE	7	12	1	-	1	-	1	1	-
	Tu'i Malila	-	-	-	8	4	18	14	14	14
<i>Total</i>		<i>49</i>	<i>54</i>	<i>104</i>	<i>12</i>	<i>10</i>	<i>19</i>	<i>69</i>	<i>82</i>	<i>100</i>

## Selection of parameters for ddRAD-seq data analysis

Within each species, assembly parameters were chosen based on a pilot study of a subset of 5 samples (*A. boucheti*: 28 samples, *A. kojimai*: 30 samples, *A. strummeri*: 46 samples). These sample subsets included triplicates of individuals for two species: six *A. boucheti* and six *A. kojimai* were included three times independently in the ddRAD-seq library preparation, starting from a unique DNA extract, therefore allowing us to estimate genotyping errors for these two species. For each pilot study, we varied one Stacks core parameter at a time (by steps of one unit: *ustack m* [3–7], *M* [2–8], and 10 *cstack n* [2–12], while the others remained at their default values, as suggested by Mastretta-Yanes et al. (2015). All analyses of these pilot studies used the following set of constant conditions: *max\_locus\_stacks* was set to 3, the minimum proportion of individuals sharing a locus was set to  $r = 0.8$ , and the minor allele count set to *MAC* = 4. The SNPs identified by Stacks were then further filtered in R version 4.1.1 (R Core Team, 2021) in order to keep the SNPs present in at least 90% of the 15 individuals and keep only the individuals which were genotyped at least at 80% of the total number of

SNPs, except in *A. strummeri*, for which we had less individuals and thus allowed a little more missing data per individual (25%).

The final set of parameters was chosen such as to maximise the number of single nucleotide polymorphism sites (SNPs) retained after filtering while maintaining a minimum genotyping error rate (measured as the frequency of differences in the three genotypes observed at each SNP and each replicate). As *A. strummeri* had no replicates in our dataset, for this species the choice of parameters was based only on the number of SNPs retained. The results of these pilot studies are presented in figures S1-3. The optimal values that minimize the genotype differences between replicates and maximize the amount of data recovered were identified as  $m=6$  (all species), and  $n=M$  set to 4, 8, and 10 for *A. boucheti*, *A. strummeri*, and *A. kojimai*, respectively. We allowed high values for  $M$  and  $n$  parameters to account for the potential occurrence of divergent haplotypes within some individuals due to introgression between divergent species, particularly in *A. strummeri* (Castel *et al.*, 2022).



30

Fig. S1 - Selection of *Stacks* parameters for *Alviniconcha boucheti*. Number of single nucleotide polymorphisms (SNPs, left panels) and genotyping error rate (right panels) after data filtering for varying values of *Stacks* parameters  $m$ ,  $M$ , and  $n$  using six *A. boucheti* individuals (3 replicates per individual).

35

Note the limited range of the Y-axis in panels C and E.

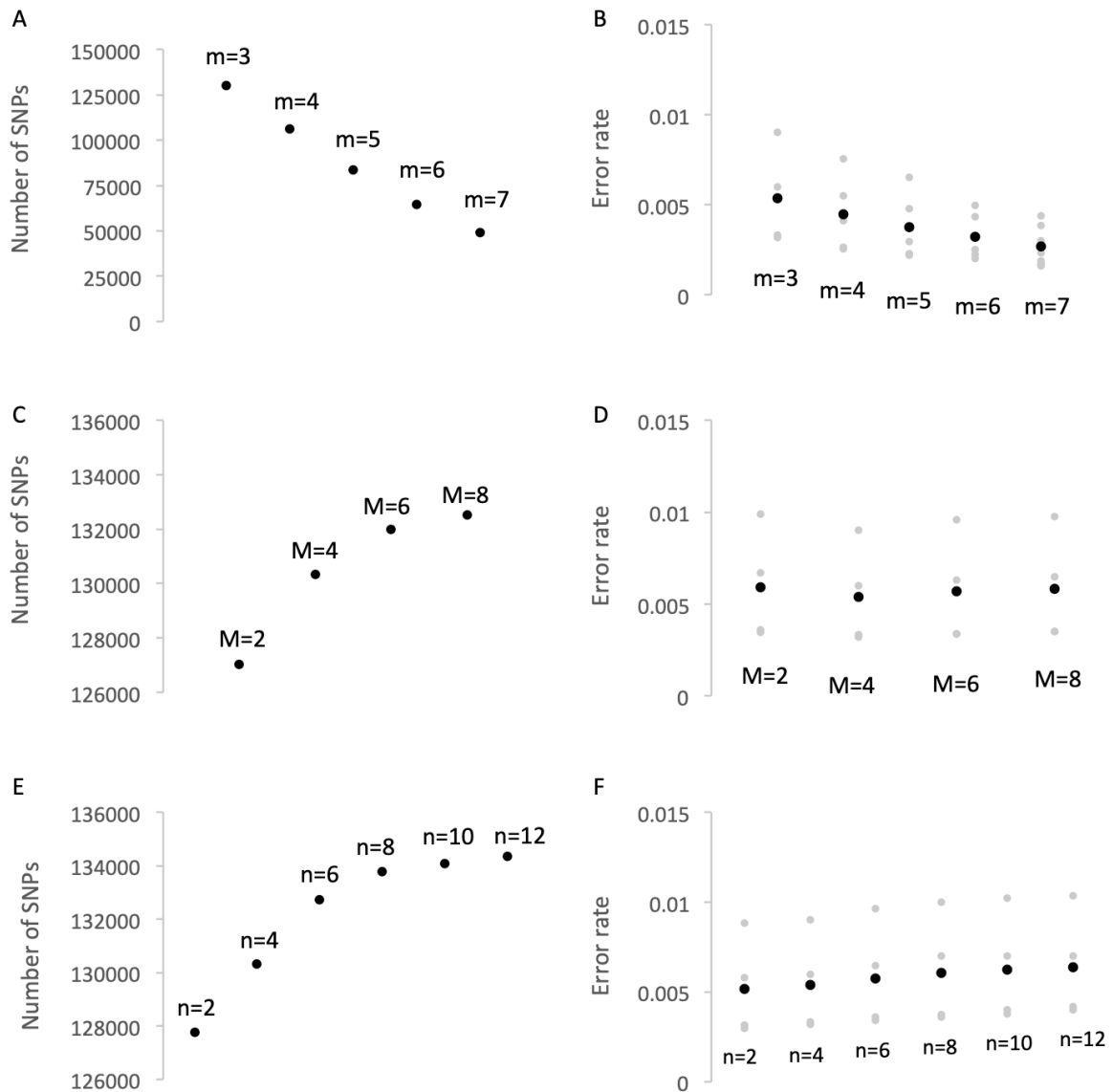
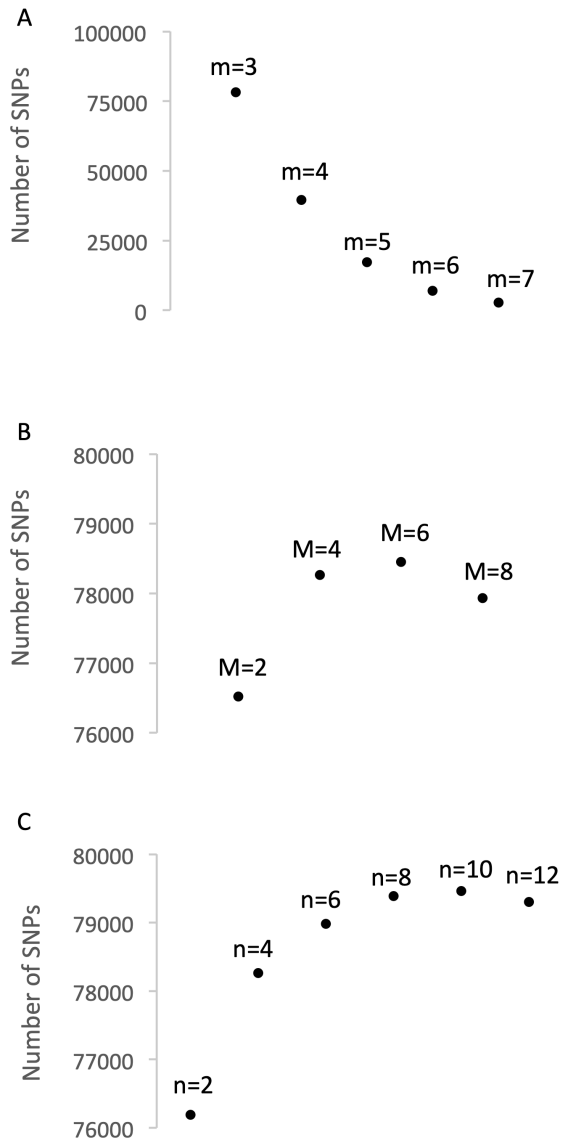
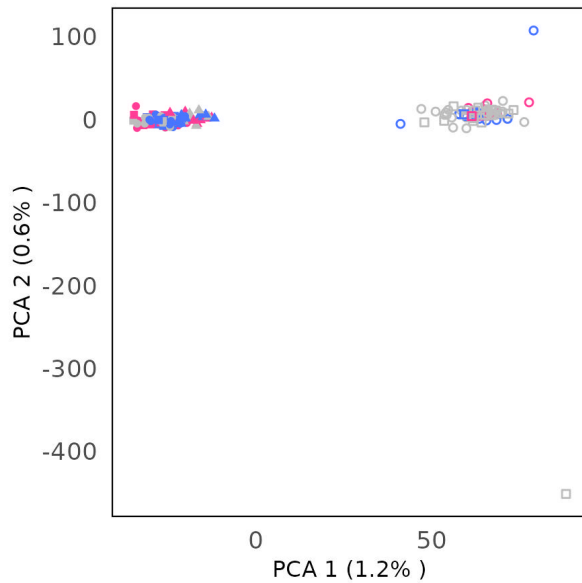


Fig. S2 - Selection of *Stacks* parameters for *Alviniconcha kojimai*. Number of single nucleotide polymorphisms (SNPs, left panels) and genotyping error rate (right panels) after data filtering for varying values of *Stacks* parameters  $m$ ,  $M$ , and  $n$  using five *A. kojimai* individuals (3 replicates per individual). Note the limited range of the Y-axis in panels C and E.



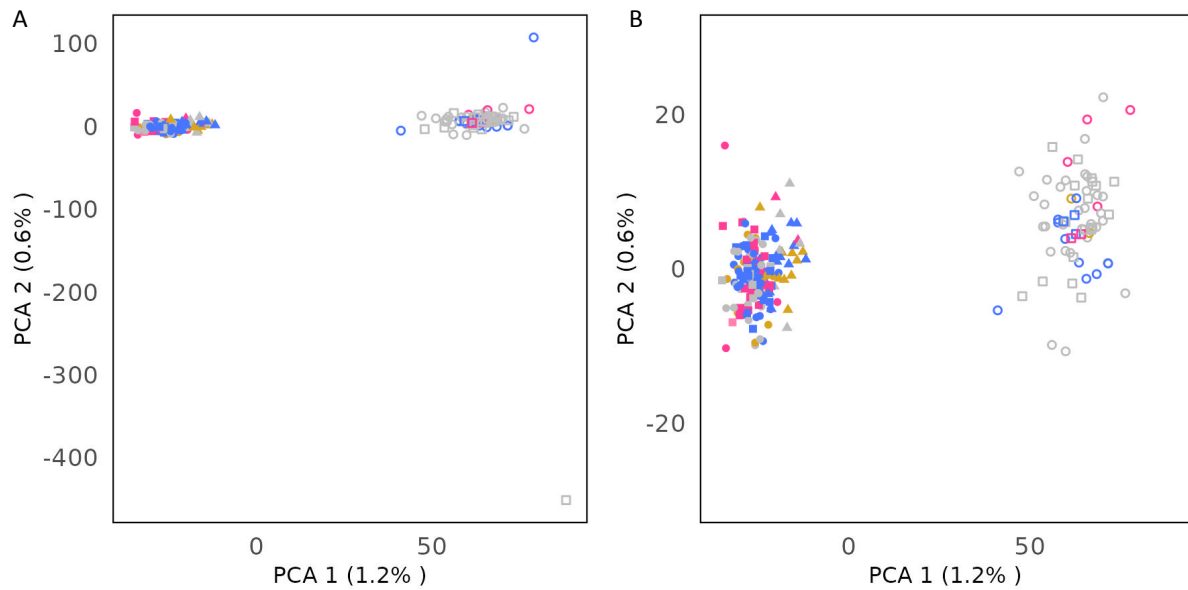
45 Fig. S3 - Selection of *Stacks* parameters for *Alviniconcha strummer*. Number of single nucleotide polymorphisms (SNPs) after data filtering for varying values of *Stacks* parameters  $m$ ,  $M$ , and  $n$ . Note the limited range of the Y-axis in panels B and C.



50 Fig. S4 - Principal components Analysis on *A. kojimai* genotypes (251 individuals, 70 122 SNPs). Axis 2 is driven strongly by a single individual (visible here at the bottom-left corner). Main text figure 2C presents the same analysis (including this outlier individual) but zoomed in to see better the distribution of all other points.

55 **Genetic analyses performed separately for males, females, and morphological hermaphrodites in *A. kojimai*.**

The analysis of genetic diversity for *A. Kojimai* (presented in main text Table 1) was repeated using three sex classes: males, females, and hermaphrodites, identified using data from the histological  
 60 analyses only. This represents 91 individuals, genotyped using 70122 SNPs. The global gene diversity  $H_T$  was 0.1115 and  $F_{ST}$  across sexes were as follows:  $F_{ST, \text{male-female}} = -0.00034$  ( $p$ -value = 0.9),  $F_{ST, \text{female-hermaphrodite}} = -0.00016$  ( $p$ -value = 0.7),  $F_{ST, \text{male-hermaphrodite}} = 0.00014$  ( $p$ -value = 0.127).



65 Figure S5: Principal components of SNP genotypes for *A. kojimai* (247 individuals, 70 122 snps). Panel  
 A) The first axis separates individuals according to their geographic origin: empty symbols correspond  
 to samples from the Manus and Woodlark basins (north-west of the study area, see main text figure  
 1), solid symbols correspond to samples from the Futuna, North Fiji and Lau areas (South-East). There  
 is no visible clustering by sex (orange symbols: morphologically androdioecious individuals, pink  
 70 symbols: individuals where only female gonadic tissue was observed, blue symbols: males). The second  
 axis is driven by two individuals. B) Close-up on the main clouds of dots from panel A.

Table S2 - Diversity indices for *A. kojimai* using  
 histological analyses.  $H_T$  : total gene  
 diversity.  $H_S$  : sex-specific gene diversity.  $H_O$  :  
 observed heterozygosity, and  $F_{IS}$ : heterozygosity  
 deficiency.

Sex	$n$	$H_S$	$H_O$	$F_{IS}$
<i>Males</i>	49	0.1119	0.1106	0.0114
<i>Females</i>	11	0.1118	0.1098	0.0177
<i>Hermaphrodites</i>	31	0.1108	0.1093	0.0136



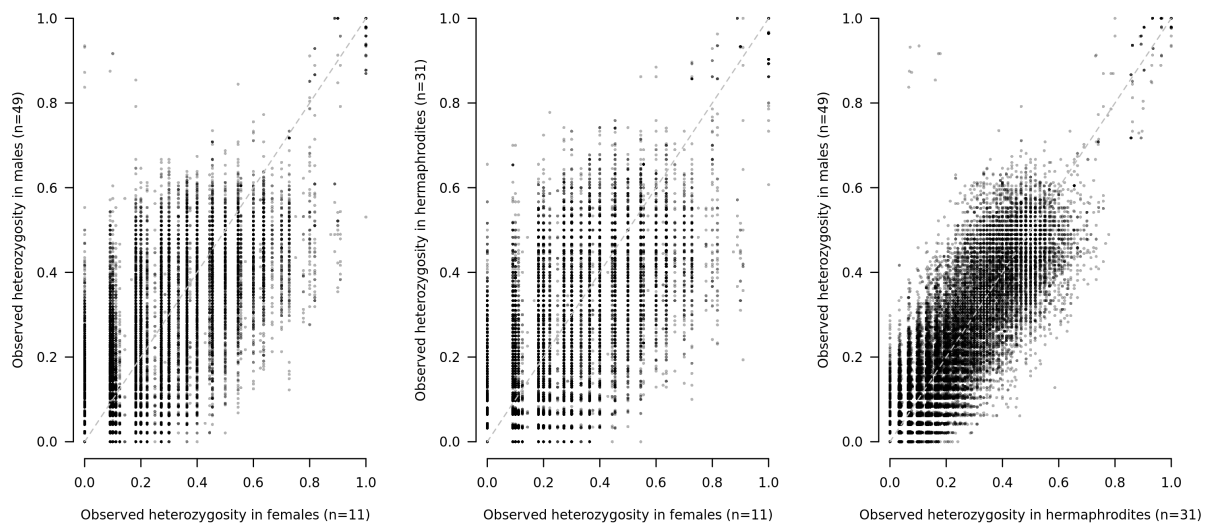


Figure S6 - Locus-specific observed heterozygosity  $H_o$  at 20 177 SNPs in *Alviniconcha kojimai* when  
 75 considering three sex classes: males ( $n=49$ ), females ( $n=11$ ), and morphological hermaphrodites  
 ( $n=31$ ). In the top-left corner of panels A and C, a group of eight SNPs consistently showed strong  
 heterozygosity in males but low heterozygosity in both hermaphrodites and females, and two  
 additional SNPs showed strong heterozygosity in males but low heterozygosity in either  
 hermaphrodites or females.

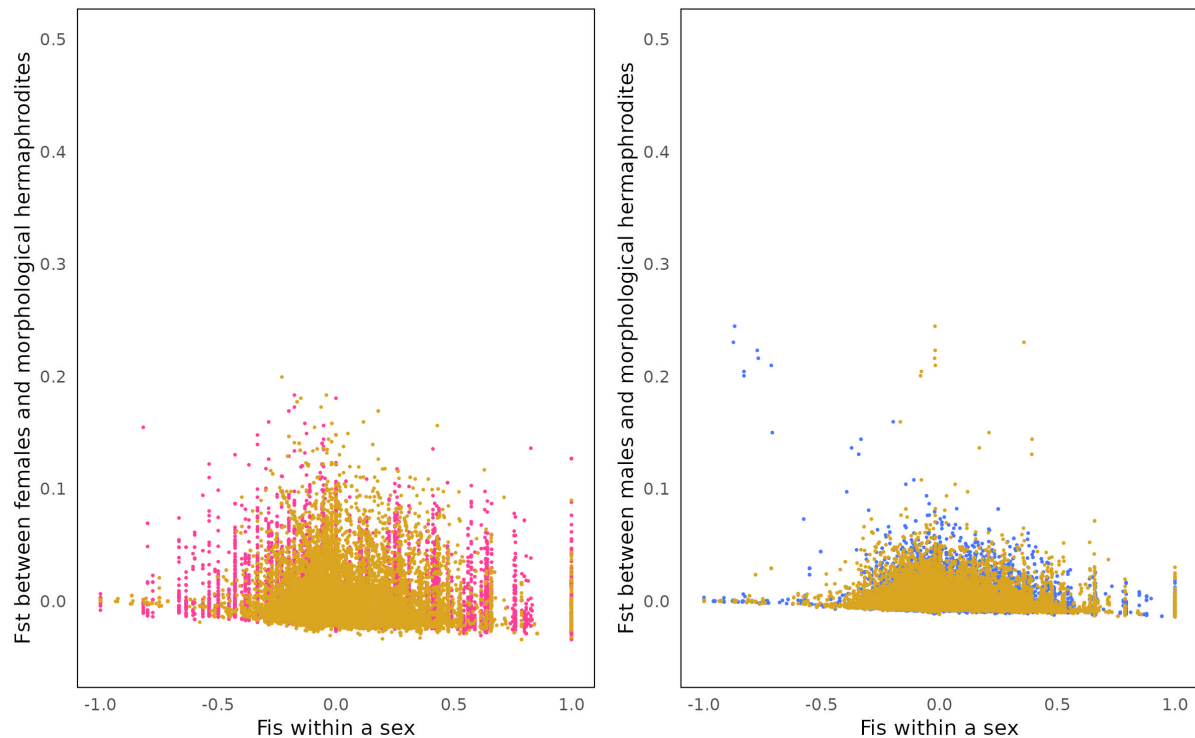


Figure S7 - Locus-specific  $F_{ST}$  between males and females and sex-specific  $F_{IS}$  for *Alviniconcha kojimai* when considering three sex classes: males ( $n=49$ , blue), females ( $n=11$ , pink), and morphological hermaphrodites ( $n=31$ , orange). Each dot shows observed values for a sex-specific  $F_{IS}$  and  $F_{ST}$  between two sex classes. The blue dots with strongest  $F_{ST}$  between males and morphological hermaphrodites and negative male  $F_{IS}$  are potential sex-linked SNPs with contrasted allelic frequencies on X and Y chromosomes in a system where the XX genotype triggers the development of the female or morphological hermaphrodite phenotype while XY triggers the development of the male phenotype (see main text).

## References

- 95 Castel, J., Hourdez, S., Pradillon, F., Daguin-Thiébaud, C., Ballenghien, M., Ruault, S., *et al.* 2022. Inter-Specific Genetic Exchange Despite Strong Divergence in Deep-Sea Hydrothermal Vent Gastropods of the Genus *Alviniconcha*. *Genes* **13**: 985. Multidisciplinary Digital Publishing Institute.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. & Emerson, B.C. 2015.

100 Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* **15**: 28–41.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing., Vienna, Austria.

105 Yang, Y., Sun, J., Chen, C., Zhou, Y., Van Dover, C.L., Wang, C., *et al.* 2022. Metagenomic and metatranscriptomic analyses reveal minor-yet-crucial roles of gut microbiome in deep-sea hydrothermal vent snail. *Animal Microbiome* **4**: 3.