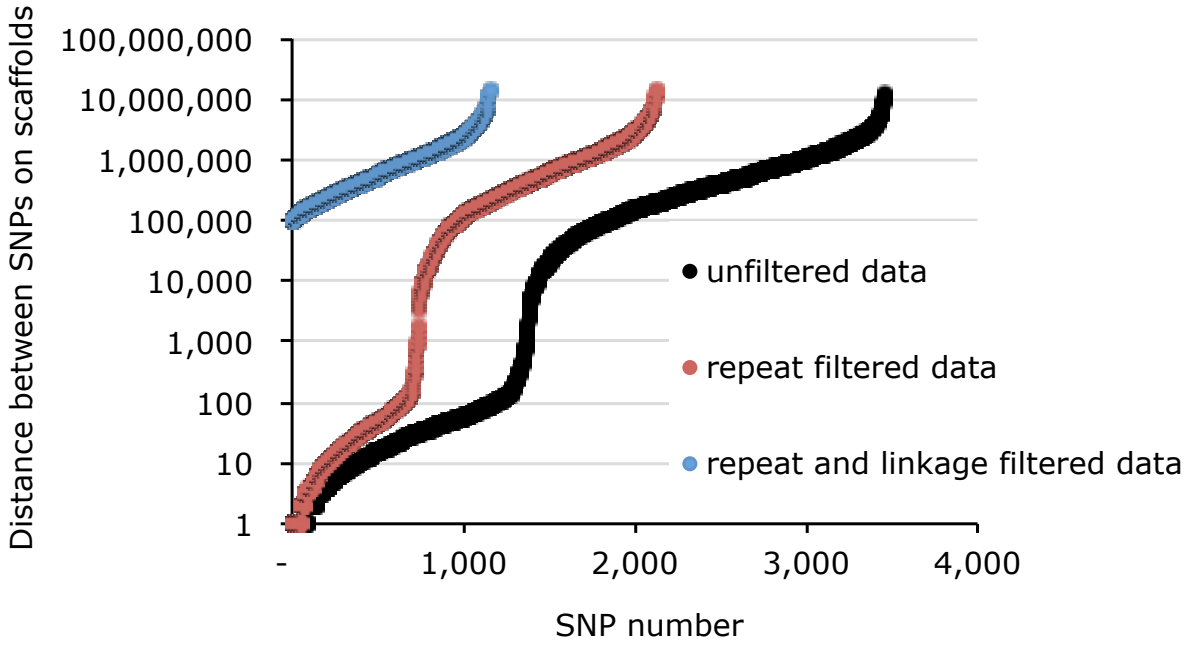
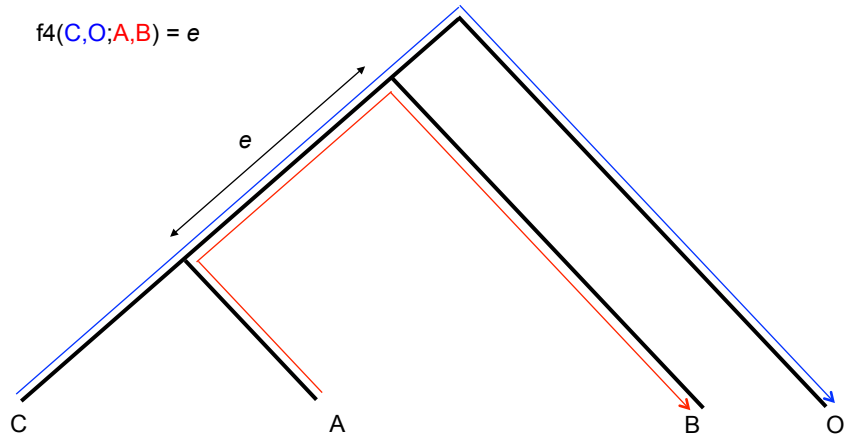
Supplementary Figure 1. A plot of the spacing between SNPs before and after each filtering step.

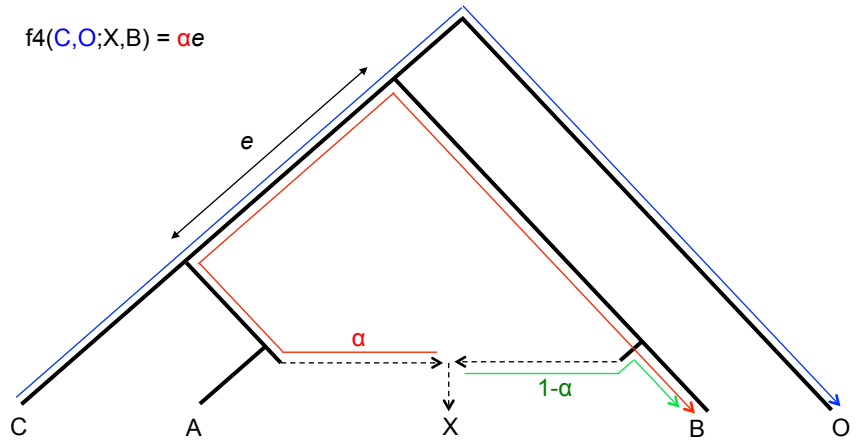


The *x*-axis is the cumulative number of SNPs, ordered based on their spacing with the next nearest SNP in the dataset. The *y*-axis is the distance in base pairs (bp) to the next nearest SNP on the same scaffold. The black markers represent the unfiltered data, constituting the 3,678 SNPs in the VCF file generated by Moura *et al*. (2014a), accessed from the Dryad data depository (doi: 10.5061/dryad.qk22t). The red markers represent the same data after SNPs that fall into regions of known repetitive elements, excessive or low coverage, or low mappability, or putatively selected loci have been filtered out. The blue line represents the data after further filtering to randomly remove one SNP from pairs of SNPs that are <100,000 bp apart on the same scaffold, resulting in a dataset of 1,346 putatively unlinked, neutrally evolving SNPs.

Supplementary Figure 2. f_4-ratio estimation of admixture proportions.



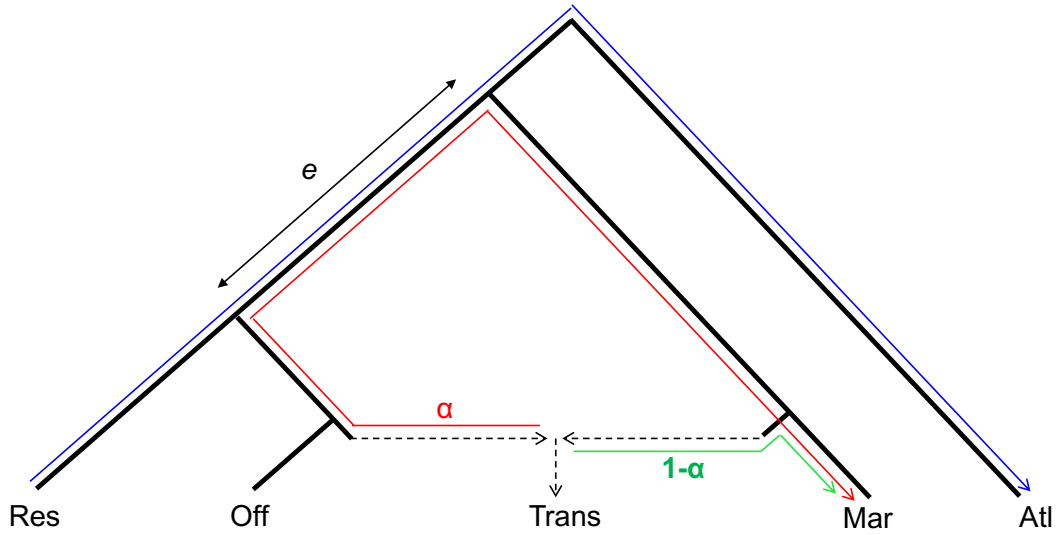
When we estimate the f_4-statistic for an incorrect topology, in this case f_4(C,O; A,B), the drift paths from C to O and from A to B overlap along edge *e*, resulting in a significant f_4-statistic:
f_4(C,O; A,B) = *e*



If we introduce admixture into this scenario, then drift can take different paths between populations. In this case, population X results from admixture between populations closely related to A and B (let us call them A´ and B´). If we consider f_4(C,O;X,B), then the drift path X to B can take a similar path to before, along edge *e*, but only with probability α, as it can also take a second path to B with probabiliy 1- α. Therefore f_4 (C,O;X,B) = α*e*

We can thus estimate the admixture proportions of X, in terms of the relative contributions of A´ and B´, using the f_4 ratio estimator of Patterson *et al*. (2012).
f_4 (C,O;X,B)/*f4*(C,O;A,B) = α

However, an important consideration that is often overlooked, is that if the admixture event predated the split time of A and C, then α can be understimated, see http://www.mailund.dk/index.php/2014/12/17/estimating-admixture-proportions/comment-page-1/



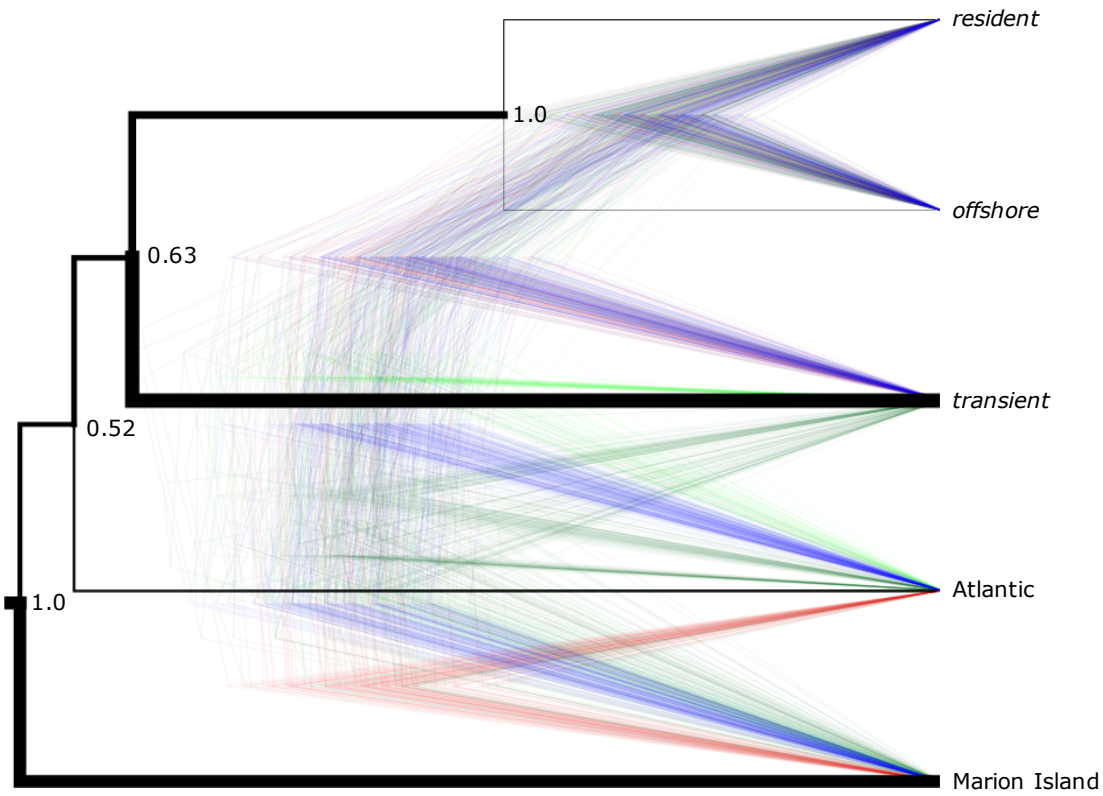
$f4(\text{Res},\text{Atl};\text{Trans},\text{Mar})/f4(\text{Res},\text{Atl};\text{Off},\text{Mar}) = \alpha$

$0.0040 / \ 0.0089 = \alpha = 0.449$

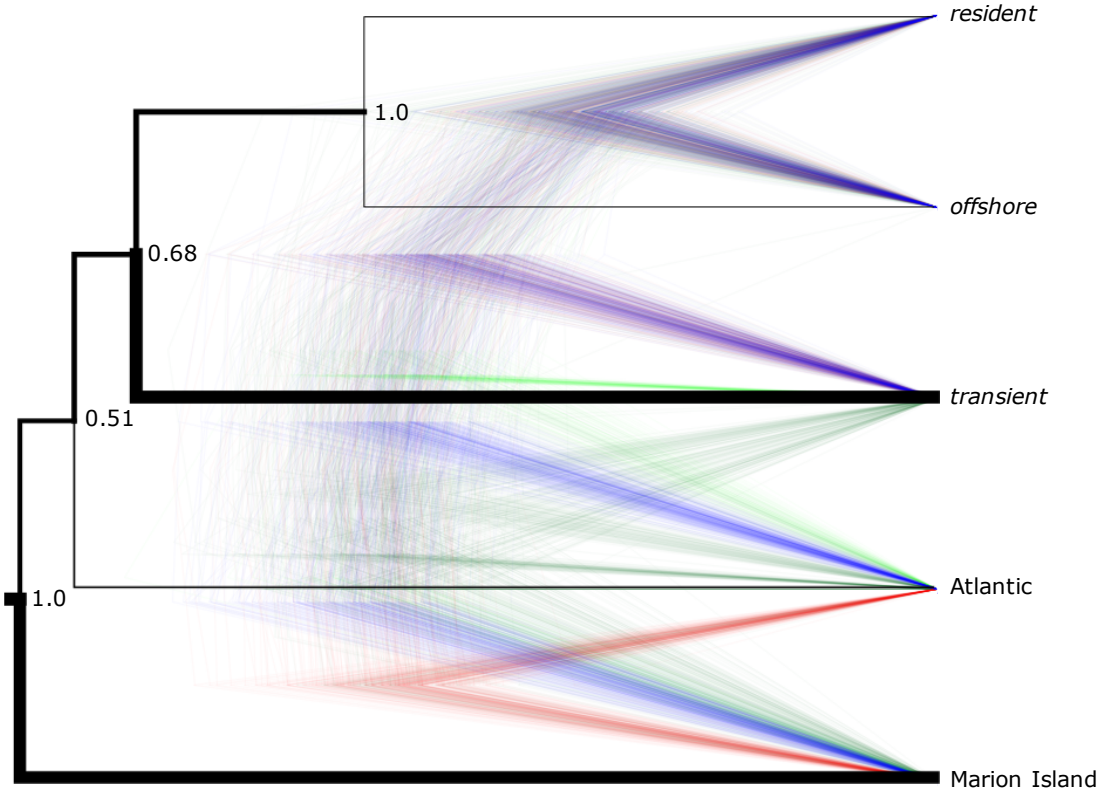
So for example, our estimate that the *transient* ecotype shares approximately 45% (based on the above estimate of α) ancestry with the *offshore* ecotype compared with 55% with the Marion Island population depends upon the admixture occurring after the *resident* and *offshore* ecotype split.

Supplementary Figure 3. Nuclear SNP phylogeny of subsamples 1-4 of different individuals as listed in Table S1.

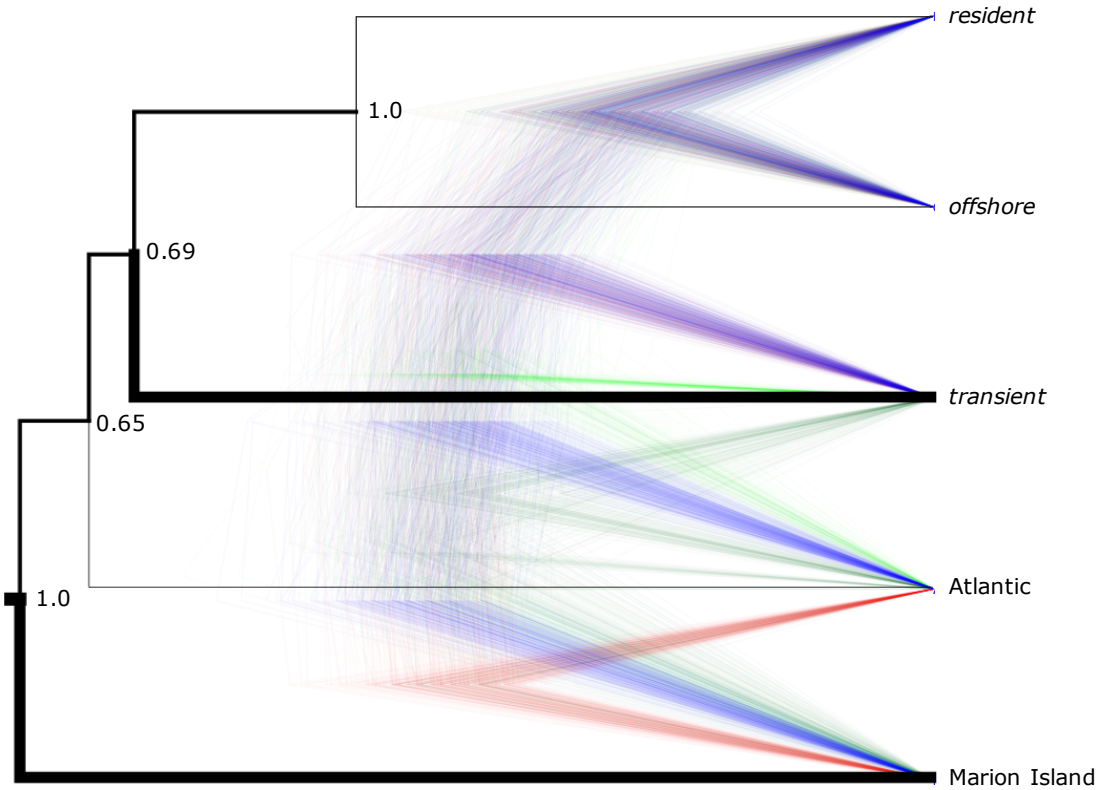
(a)



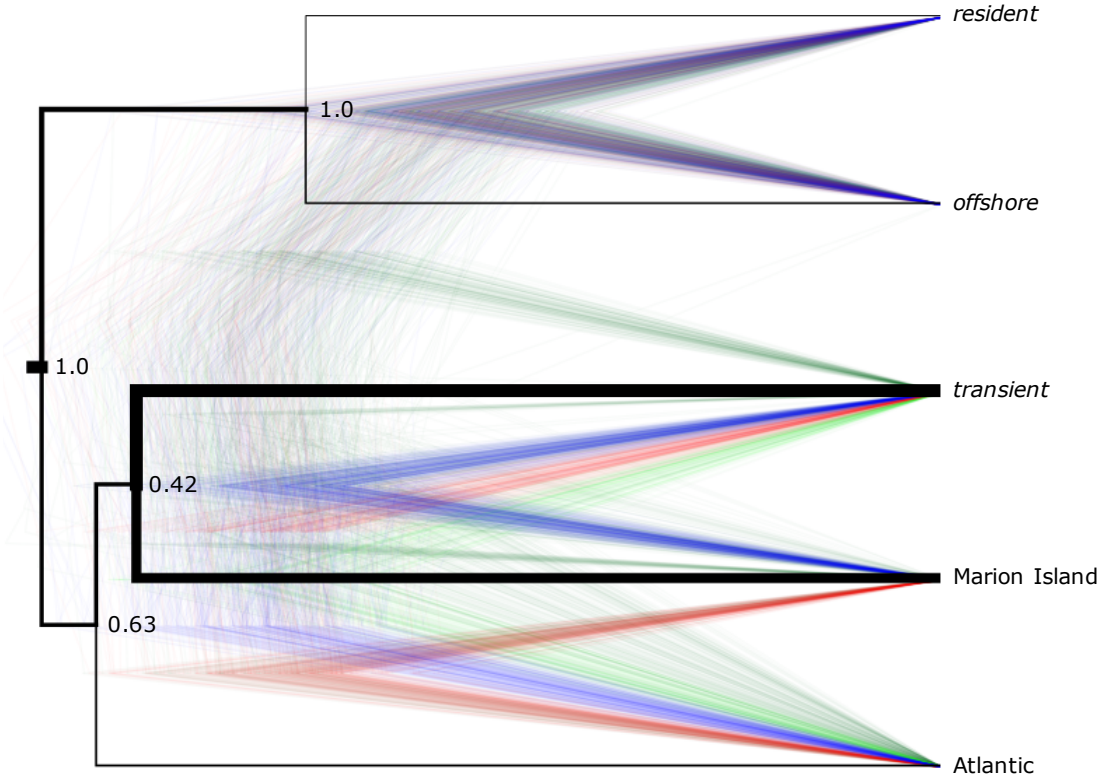
(b)



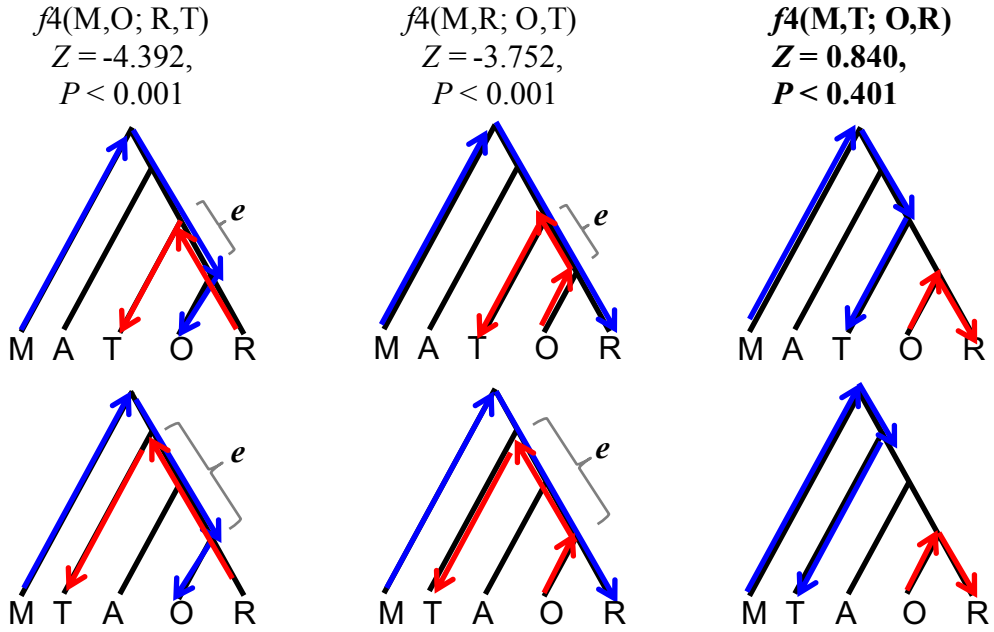
(c)


resident
offshore
transient
Atlantic
Marion Island
1.0
0.69
0.65
1.0

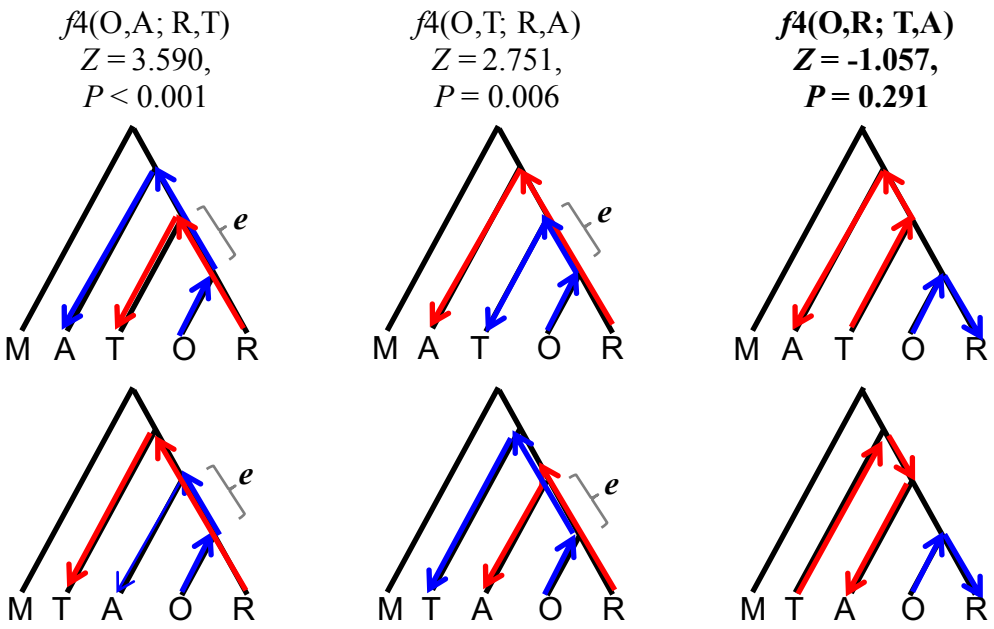
(d)


resident
offshore
transient
Marion Island
Atlantic
1.0
1.0
0.42
0.63

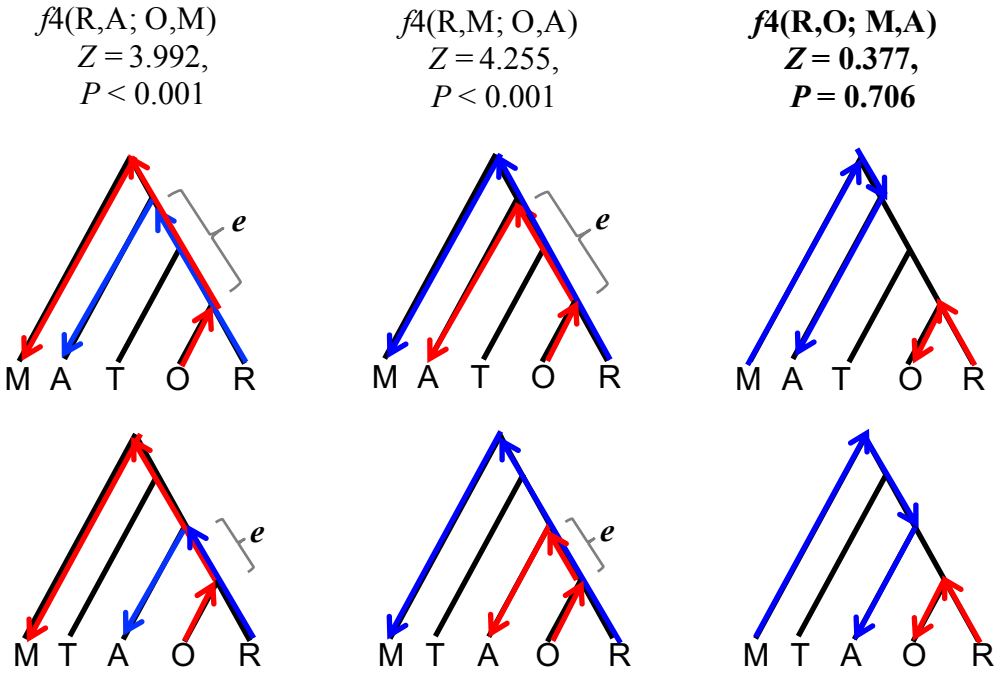
Supplementary Figure 4. A comparison of the expected (the visualised drift paths) and observed value of the f_4-statistics.



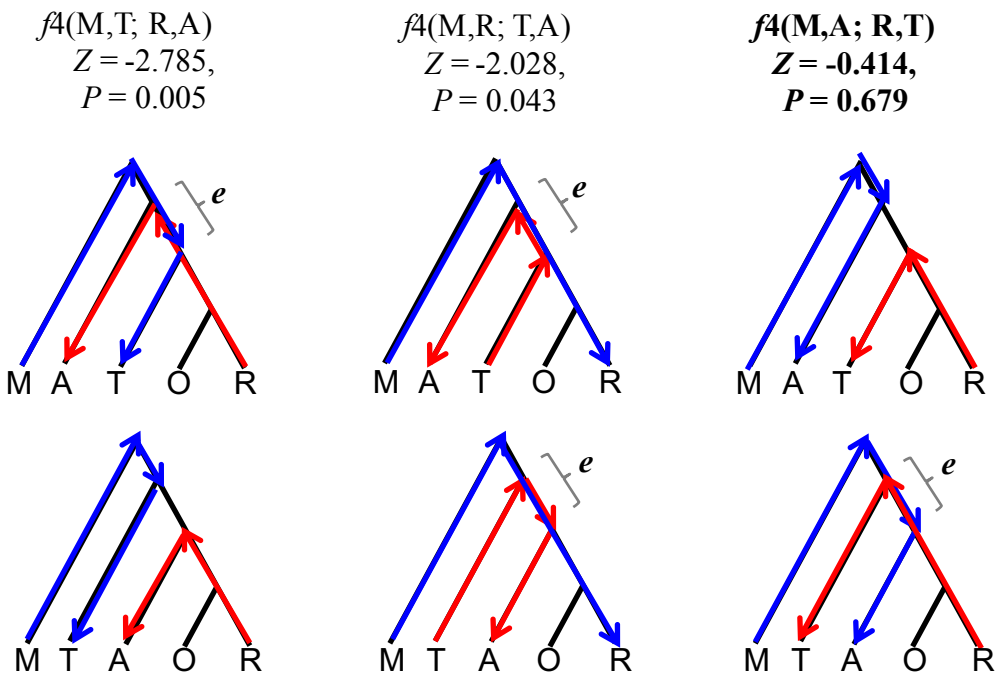
(a) A comparison of the expected (the visualised drift paths) and observed value of the f_4-statistics for the three possible tree-like relationships among the following four populations: Marion Island (M), North Pacific *Transient* (T), North Pacific *Resident* (R) and North Pacific *Offshore* (O) and assuming the topology reconstructed from allele frequencies of SNPs generated by SNAPP (top row) and the topology inferred from concatenated RAD sequences by Moura *et al*., (2015) (bottom row). In this example the observed and expected f_4-statistic of the form f_4(A,B; C,D) are in agreement, as there is no significant observed covariance in drift between (M,T) and (O,R) as expected and the tree (M,T; O,R) appears to describe the relationship among these populations, but the trees (M,O; R,T) and (M,R; O,T) are rejected as the drift paths overlap along edge *e*, and the observed f_4-statistic is significantly different from zero. However, we note that although drift paths overlap along the same edge *e* for trees (M,O; R,T) and (M,R; O,T), we observe some difference in the f_4- and Z-statistics, suggesting that some drift is taking another path in (M,R; O,T). The f_4-statistic estimated higher rates of drift along edges estimated by SNAPP as having low theta, as would be expected in populations with lower effective population size (theta is scaled by effective population size). Note that statistical tests within each combination of taxa, and across different combinations of taxa are not independent of each other but should be viewed as the strength of support for deviations from a tree-like model of population branching across various four-taxon subsets.



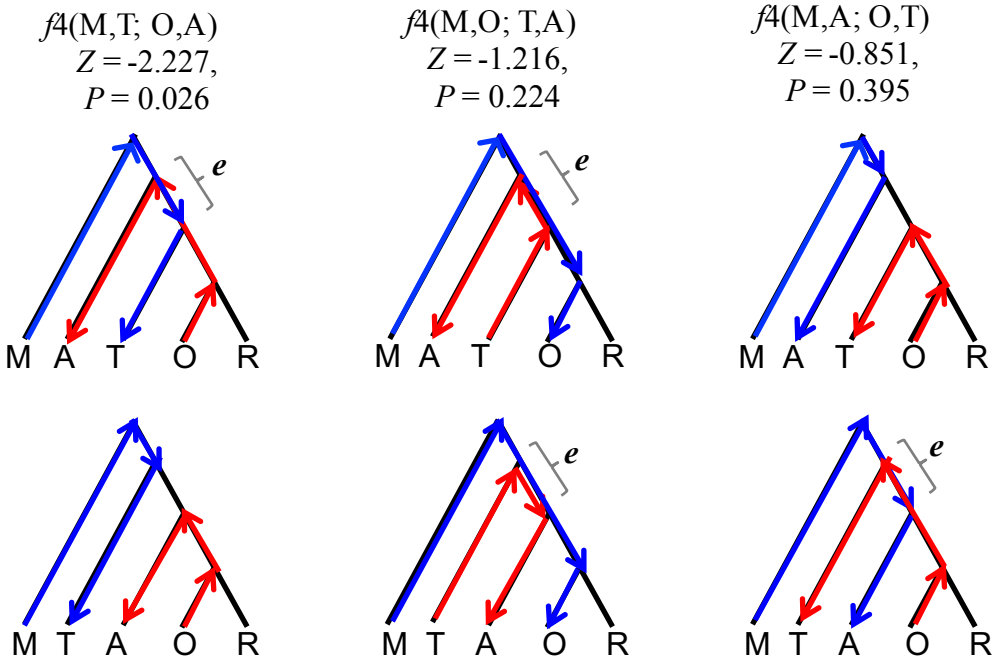
(b) As in figure **4a**, here we present a comparison of the expected and observed value of the f_4-statistics for the three possible tree-like relationships in this example among the populations: Atlantic (A), North Pacific *Transient* (T), North Pacific *Resident* (R) and North Pacific *Offshore* (O). Again, the observed and expected f_4-statistics are in agreement. The tree (O,R; T,A) appears to describe the relationship among these populations, but the trees (O,A; R,T) and (O,T; R,A) are rejected. In contrast to the previous example, here the drift paths are in the same direction along edge *e*, hence the positive f_4-statistic and Z-score.



(c) A comparison of the expected and observed value of the f_4-statistics for the three possible tree-like relationships among the two outgroups Atlantic (A) and Marion Island (M), and the North Pacific *Resident* (R) and *Offshore* (O) ecotypes. Again, the observed and expected f_4-statistics are in agreement. The tree (R,O; M,A) appears to describe the relationship among these populations.

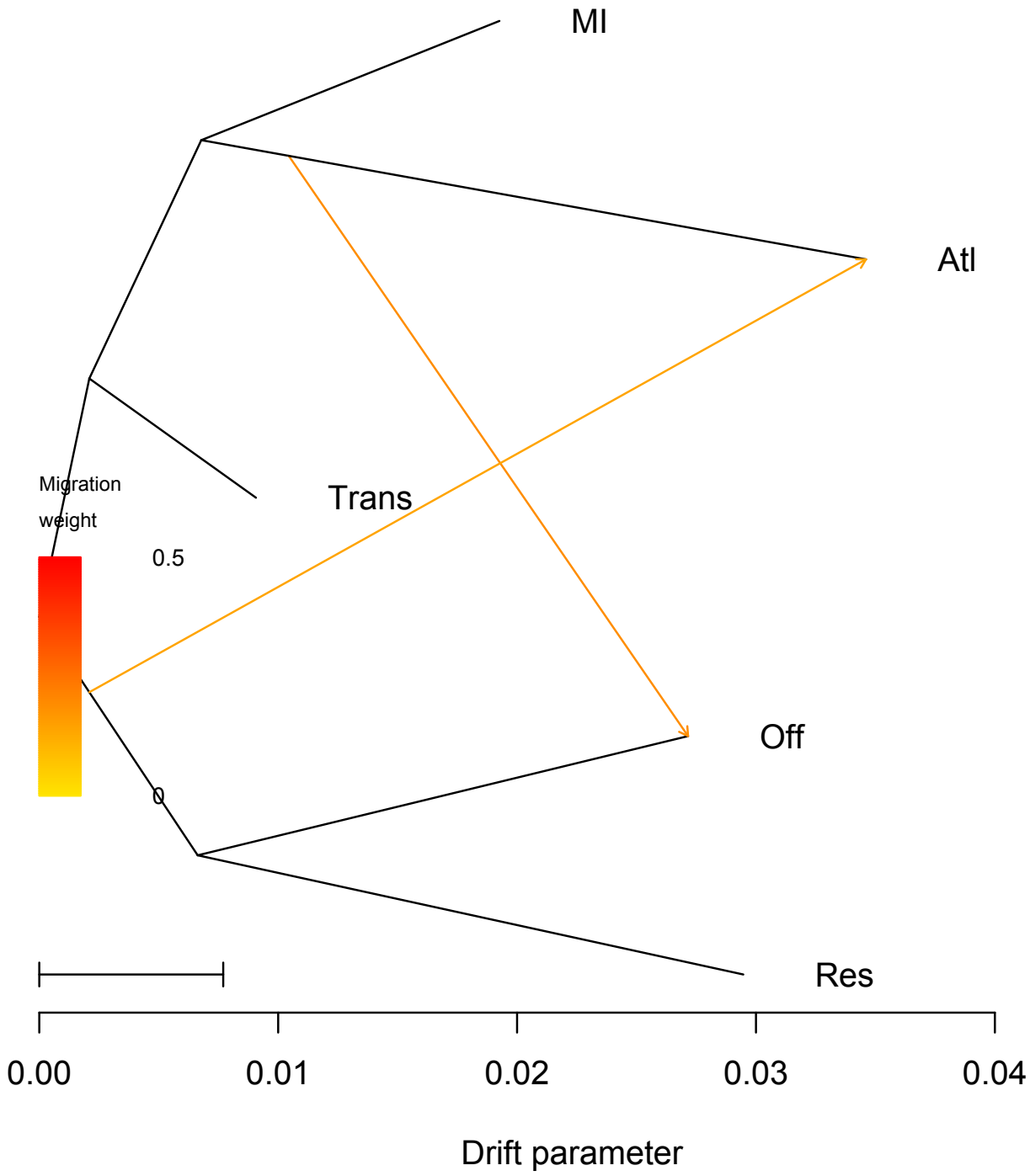


(d) (M,A; R,T) appears to describe the relationship among these populations. The expected and observed f_4-statistics agree when the topology inferred using SNAPP is assumed, but not when the topology from Moura et al.'s concatenated RAD-seq data is assumed.

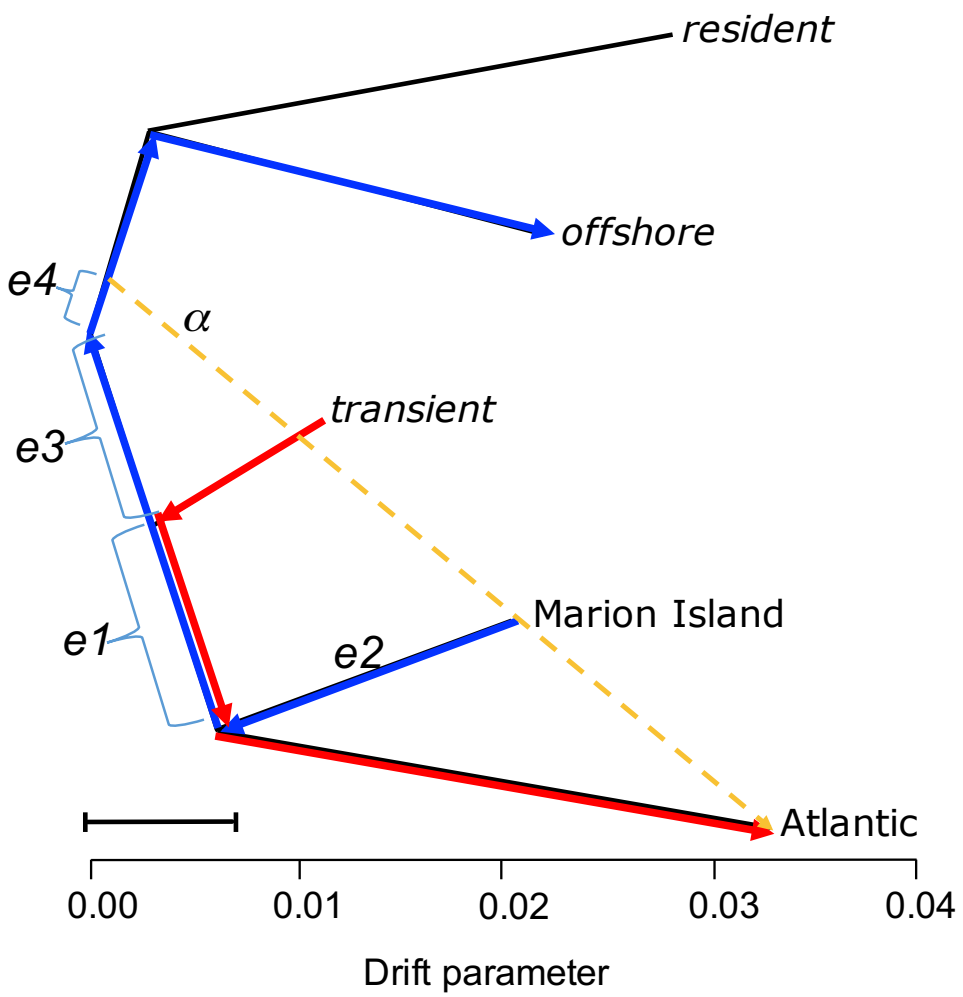


(e) In this example we find no clear support from the observed data for any one of the three possible trees that could describe the relationships among the four populations (M, T, A and R) above the other two trees. The observed and expected f_4-statistic are discordant and the relationship among these four populations does not appear to be fully described by a simple tree model.

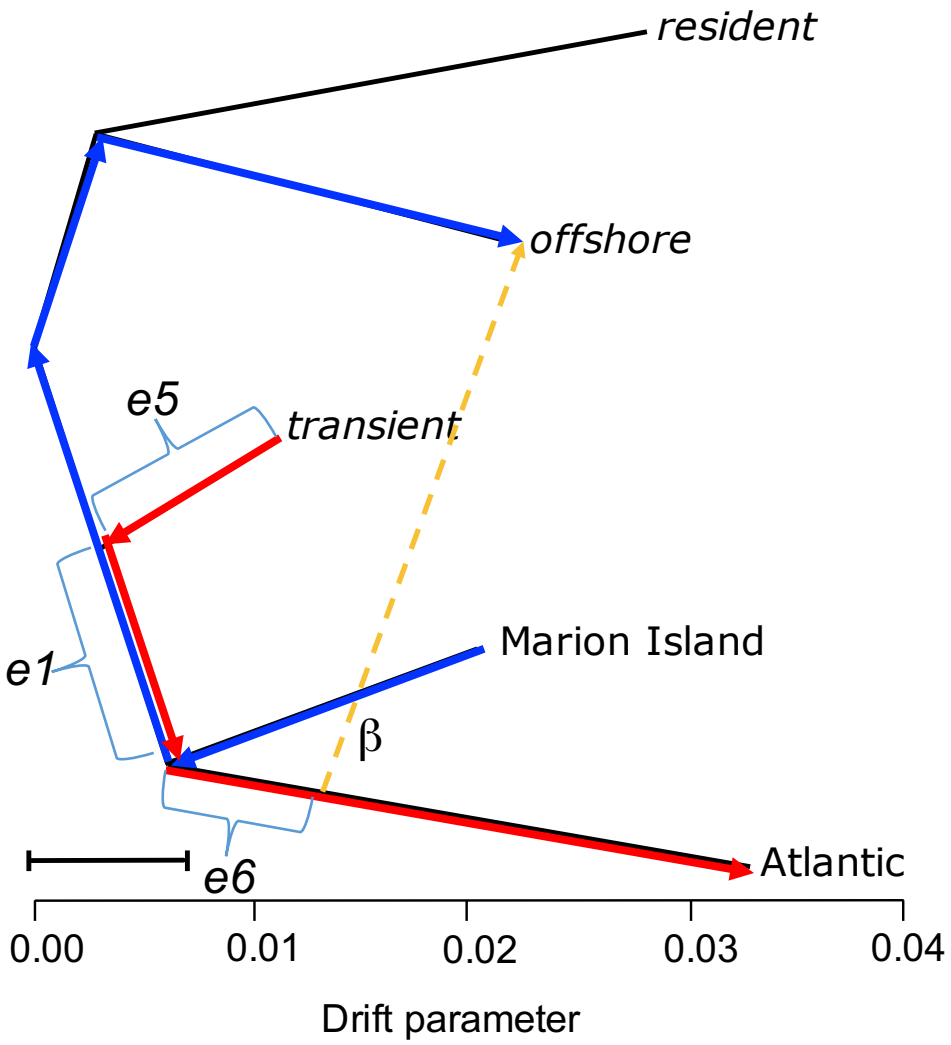
Supplementary Figure 5. TreeMix inference of admixture events.



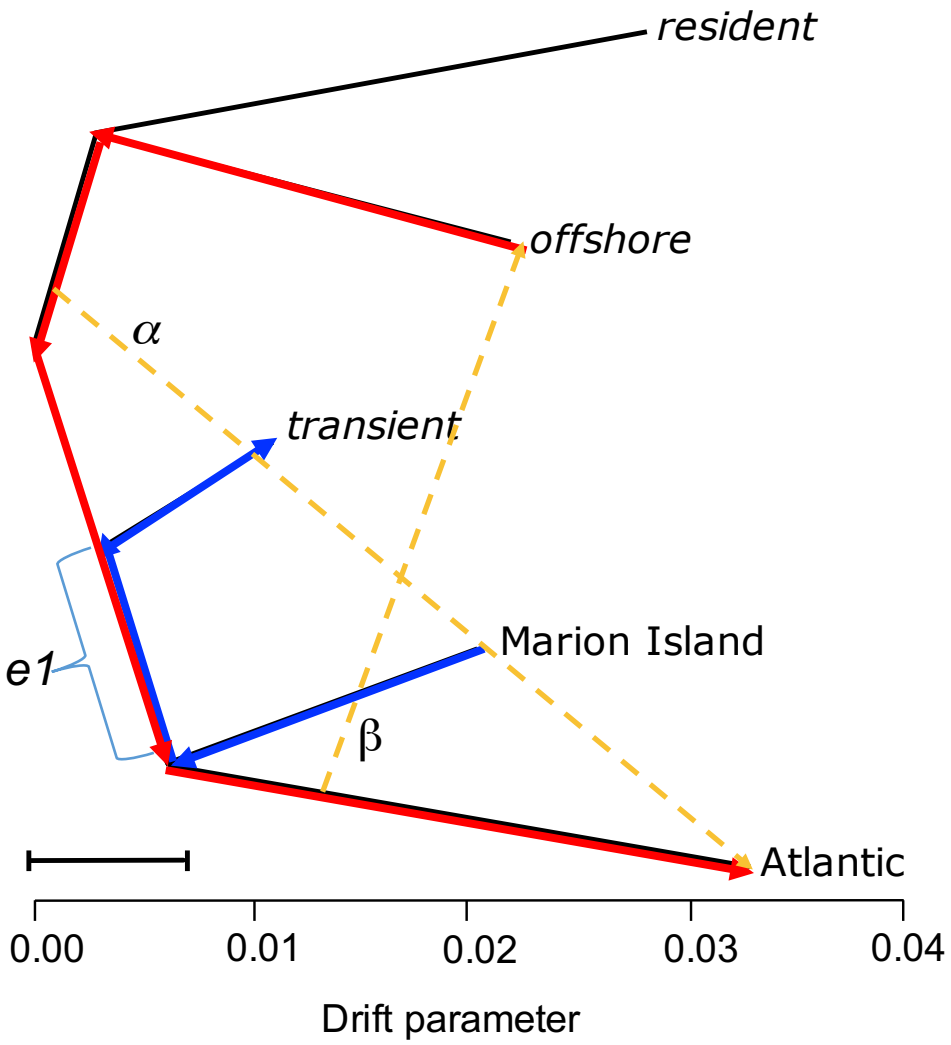
(a) TreeMix admixture graph visualising the relationship among populations as a bifurcating maximum-likelihood tree, with migration edges between populations that are a poor fit for this tree model. The strength and directionality of the inferred introgression is indicated by the coloured arrows. The scale bar represents 10 times the average standard error (s.e.) of the values in the covariance matrix.



(b) The migration edges inferred by TreeMix and visualised above in Supplementary Figure 5a, are added to the tree from Figure 3a for simplicity, and the drift paths from the f_4 statistic test f_4(M,O; T,A) are coloured in blue and red. Although the f_4 statistic is expected to be significant and negative due to drift taking overlapping paths in the opposite directions along edge *e*1, we see that the migration events reduce the extent of this covariance in allele frequency changes.
For example, if we consider the blue drift path from Marion Island to *offshore*, we see that at least some of the overlapping drift that occurs along path *e*1 in the opposing direction to the red drift path from *transient* to Atlantic will be cancelled out, as the drift from Marion Island to *offshore* along edges *e1-e4* will be shared via the migration edge with the Atlantic population with probability α, where α depends upon the weight of migration.



(c) Similarly, if we consider the red drift path from *transient* to Atlantic, we see that some of the overlapping drift that occurs along path *e*1 in the opposing direction to the blue drift path from Marion Island to *offshore* will be cancelled out, as the drift from *transient* to Atlantic along edges *e1*, *e5* & *e6* will be shared via the migration edge with the *offshore* ecotype with probability β, where β depends upon the weight of migration. Thus, f_4(M,O; T,A) = *e1*-(α(*e1*+*e2*+*e3*+*e4*)+ β(*e1*+*e5*+*e6*)).



(d) In contrast to the example in (b & c), when we consider the f_4 statistic test f_4(M,T; O,A), we see that although the migration edges will reduce drift from *offshore* to Atlantic, there is no effect on the drift path from Marion Island to *transient*.

Table S1. Individual IDs of samples included in each run of SNAPP (see Figure 1 and Supplementary Figure 3). The total sample number for each ecotype/population from which these subsamples were selected are given in parentheses in the table header.

	resident (52)	***offshore*** (7)	***Transient*** (37)	**Atlantic** (6)	**Marion Island** (13)
subsample **1**	1, 16, 104, 116, 31	26, 96, 98, 100, 102	38, 48, 68, 79, 130	82, 86, 87, 91, 92	131, 133, 136, 140, 145
subsample **2**	6, 12, 22, 111, 123	26, 96, 97, 99, 102	40, 42, 58, 62, 76	82, 84, 87, 91, 92	137, 139, 141, 142, 143
subsample **3**	9, 29, 109, 113, 129	97, 98, 99, 100, 102	44, 55, 56, 70, 73	82, 84, 86, 91, 92	132, 134, 135, 140, 145
subsample **4**	11, 15, 24, 108, 120	96, 97, 98, 99, 100	46, 53, 59, 64, 75	82, 84, 86, 87, 92	132, 133, 136, 137, 143
subsample **5**	6, 12, 22, 111, 123 9, 29, 109, 113, 129	26, 96, 97, 98, 99, 100, 102	40, 42, 58, 62, 76 44, 55, 56, 70, 73	82, 84, 86, 87 91, 92	132, 134, 135, 137, 139 140, 141, 142, 143, 145

Table S2. Estimates of theta and 95% highest posterior density interval (HPDI) for each branch and each run of SNAPP.

	Estimates of theta (95% HPDI)				
	subsample 1	**subsample 2**	**subsample 3**	**subsample 4**	**subsample 5**
resident	0.051 (0.038-0.067)	0.052 (0.038-0.068)	0.062 (0.045-0.080)	0.056 (0.045-0.077)	0.040 (0.030-0.052)
offshore	0.058 (0.042-0.074)	0.072 (0.056-0.094)	0.081 (0.057-0.104)	0.093 (0.072-0.119)	0.055 (0.040-0.070)
transient	0.273 (0.219-0.327)	0.267 (0.211-0.319)	0.244 (0.193-0.299)	0.269 (0.209-0.319)	0.219 (0.178-0.263)
Atlantic	0.091 (0.072-0.106)	0.084 (0.067-0.104)	0.077 (0.060-0.093)	0.096 (0.078-0.115)	0.079 (0.064-0.094)
Marion Island	0.246 (0.198-0.298)	0.257 (0.211-0.317)	0.249 (0.199-0.302)	0.221 (0.175-0.262)	0.225 (0.189-0.263)
Anc1	0.126 (0.066-0.189)	0.132 (0.057-0.204)	0.128 (0.069-0.190)	0.127 (0.063-0.196)	0.126 (0.065-0.196)
Anc2	0.138 (0.074-0.217)	0.142 (0.073-0.230)	0.129 (0.060-0.199)	0.128 (0.065-0.198)	0.170 (0.084-0.279)
Anc3	0.171 (0.110-0.237)	0.147 (0.082-0.226)	0.129 (0.073-0.199)	0.147 (0.073-0.222)	0.145 (0.085-0.215)

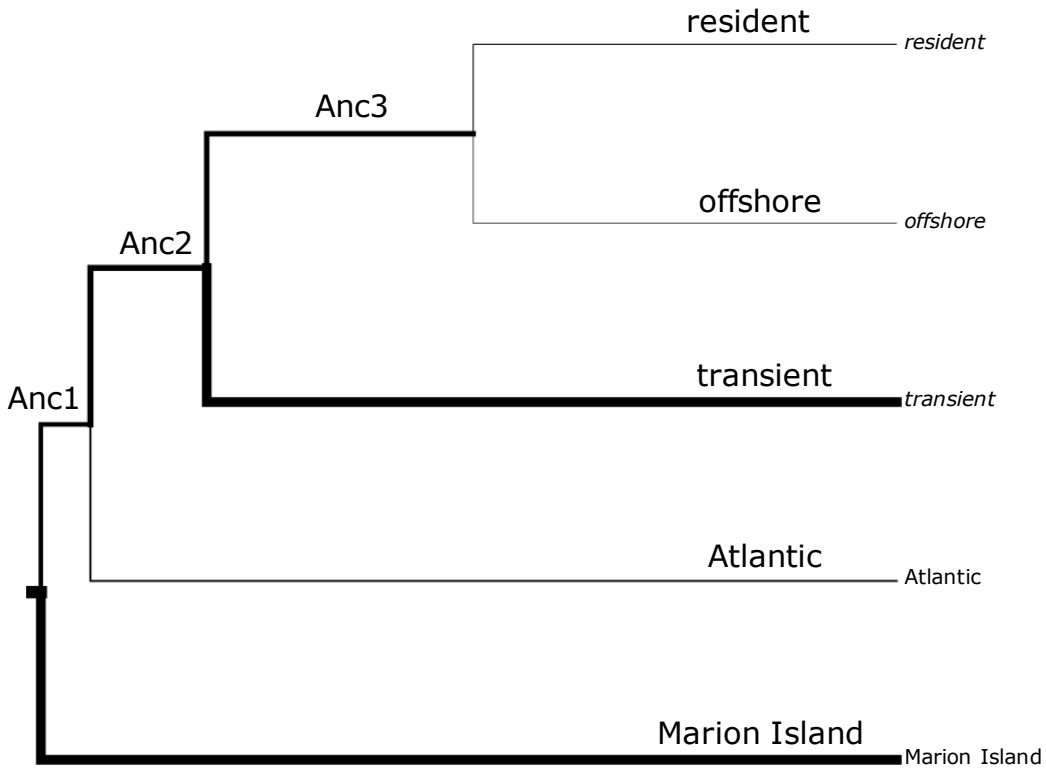


Table S3. Topologies from each run of SNAPP using different subsamples found within the 95% HPD of each using TreeSetAnalyser. The topology inferred from concatenated RAD-seq data in Moura *et al.* 2015 is highlighted in bold for each subsample.

Subsample 1

95% HPD contains 9 topologies, out of a total of 15 topologies

#nr	coverage	tree
Tree 1:	36%	((((Res,Off),Trans),Atl),Mar)
Tree 2:	18%	(((Res,Off),Trans),(Atl,Mar))
Tree 3:	**10%**	**((((Res,Off),Atl),Trans),Mar)**
Tree 4:	9%	((((Res,Off),Trans),Mar),Atl)
Tree 5:	6%	(((Res,Off),(Trans,Atl)),Mar)
Tree 6:	6%	((Res,Off),(Trans,(Atl,Mar)))
Tree 7:	5%	((Res,Off),((Trans,Mar),Atl))
Tree 8:	4%	((Res,Off),((Trans,Atl),Mar))
Tree 9:	2%	(((Res,Off),Atl),(Trans,Mar))

Subsample 2

95% HPD contains 9 topologies, out of a total of 16 topologies

#nr	coverage	tree
Tree 1:	36%	((((Res,Off),Trans),Atl),Mar)
Tree 2:	23%	(((Res,Off),Trans),(Atl,Mar))
Tree 3:	**9%**	**((((Res,Off),Atl),Trans),Mar)**
Tree 4:	9%	((((Res,Off),Trans),Mar),Atl)
Tree 5:	6%	(((Res,Off),(Trans,Atl)),Mar)
Tree 6:	5%	((Res,Off),(Trans,(Atl,Mar)))
Tree 7:	4%	((Res,Off),((Trans,Mar),Atl))
Tree 8:	3%	(((Res,Off),Atl),(Trans,Mar))
Tree 9:	2%	((Res,Off),((Trans,Atl),Mar))

Subsample 3

95% HPD contains 7 topologies, out of a total of 13 topologies

#nr	coverage	tree
Tree 1:	43%	((((Res,Off),Trans),Atl),Mar)
Tree 2:	22%	(((Res,Off),Trans),(Atl,Mar))
Tree 3:	**11%**	**((((Res,Off),Atl),Trans),Mar)**
Tree 4:	10%	(((Res,Off),(Trans,Atl)),Mar)
Tree 5:	4%	((Res,Off),(Trans,(Atl,Mar)))
Tree 6:	3%	((((Res,Off),Trans),Mar),Atl)
Tree 7:	2%	((Res,Off),((Trans,Atl),Mar))

Subsample 4

95% HPD contains 7 topologies, out of a total of 14 topologies

#nr	coverage	tree
Tree 1:	25%	((((Res,Off),Trans),Atl),Mar)
Tree 2:	21%	(((Res,Off),Trans),(Atl,Mar))
Tree 3:	10%	((Res,Off),((Trans,Mar),Atl))
Tree 4:	9%	((((Res,Off),Trans),Mar),Atl)
Tree 5:	7%	((Res,Off),(Trans,(Atl,Mar)))
Tree 6:	7%	(((Res,Off),Atl),(Trans,Mar))
Tree 7:	**6%**	**((((Res,Off),Atl),Trans),Mar)**
Tree 8:	4%	(((Res,Off),(Trans,Mar)),Atl)
Tree 9:	4%	(((Res,Off),(Trans,Atl)),Mar)
Tree 10:	4%	((Res,Off),((Trans,Atl),Mar))

Subsample 4

95% HPD contains 6 topologies, out of a total of 8 topologies

#nr	coverage	tree
Tree 1:	31%	((Res,Off),((Trans,Mar),Atl))
Tree 2:	24%	((Res,Off),(Trans,(Atl,Mar)))
Tree 3:	19%	(Res,(Off,((Trans,Mar),Atl)))
Tree 4:	10%	((Res,Off),((Trans,Atl),Mar))
Tree 5:	7%	(Res,(Off,(Trans,(Atl,Mar))))
Tree 6:	6%	(Res,(Off,((Trans,Atl),Mar)))

Subsample 5

95% HPD contains 3 topologies, out of a total of 10 topologies

#nr	coverage	tree
Tree 1:	61%	((((Res,Off),Trans),Atl),Mar)
Tree 2:	23%	(((Res,Off),Trans),(Atl,Mar))
Tree 3:	13%	((((Res,Off),Trans),Mar),Atl)

Table S4. f_4 statistic results for comparisons of different topologies. Runs were done using different jackknife block sizes to estimate standard error (s.e.).

Four-taxon tree	*f_4 statistic*	*s.e.*	*Z-score*
Estimating f_4 in 461 blocks of size 5			
Res,MI;Trans,Off	-0.0061	0.0016	-3.752
Res,Off;Trans,MI	0.0009	0.0011	0.840
Res,Trans;Off,MI	0.0070	0.0016	4.392
Res,Atl;Trans,Off	-0.0049	0.0018	-2.751
Res,Off;Trans,Atl	0.0015	0.0014	1.056
Res,Trans;Off,Atl	0.0064	0.0018	3.590
Res,Atl;Trans,MI	0.0040	0.0014	2.785
Res,MI;Trans,Atl	0.0034	0.0017	2.028
Res,Trans;MI,Atl	-0.0006	0.0015	-0.414
Res,Atl;Off,MI	0.0089	0.0022	3.992
Res,MI;Off,Atl	0.0095	0.0022	4.255
Res,Off;MI,Atl	0.0006	0.0015	0.377
Trans,Atl;Off,MI	0.0019	0.0016	1.216
Trans,MI;Off,Atl	0.0031	0.0014	2.227
Trans,Off;MI,Atl	0.0012	0.0014	0.851
Estimating f_4 in 230 blocks of size 10			
Res,MI;Trans,Off	-0.0061	0.0016	-3.760
Res,Off;Trans,MI	0.0009	0.0010	0.902
Res,Trans;Off,MI	0.0070	0.0016	4.571
Res,Atl;Trans,Off	-0.0049	0.0018	-2.791
Res,Off;Trans,Atl	0.0015	0.0015	1.027
Res,Trans;Off,Atl	0.0064	0.0019	3.347
Res,Atl;Trans,MI	0.0040	0.0014	2.795
Res,MI;Trans,Atl	0.0034	0.0017	1.976
Res,Trans;MI,Atl	-0.0006	0.0015	-0.412
Res,Atl;Off,MI	0.0089	0.0022	4.124
Res,MI;Off,Atl	0.0095	0.0023	4.104
Res,Off;MI,Atl	0.0006	0.0015	0.386
Trans,Atl;Off,MI	0.0019	0.0016	1.189
Trans,MI;Off,Atl	0.0030	0.0013	2.332
Trans,Off;MI,Atl	0.0012	0.0014	0.861
Estimating f_4 in 46 blocks of size 50			
Res,MI;Trans,Off	-0.0061	0.0017	-3.599
Res,Off;Trans,MI	0.0009	0.0010	0.874
Res,Trans;Off,MI	0.0070	0.0015	4.774
Res,Atl;Trans,Off	-0.0049	0.0020	-2.495
Res,Off;Trans,Atl	0.0015	0.0015	1.010
Res,Trans;Off,Atl	0.0064	0.0020	3.191
Res,Atl;Trans,MI	0.0040	0.0015	2.601
Res,MI;Trans,Atl	0.0034	0.0019	1.759
Res,Trans;MI,Atl	-0.0006	0.0014	-0.427
Res,Atl;Off,MI	0.0089	0.0025	3.505
Res,MI;Off,Atl	0.0095	0.0025	3.754
Res,Off;MI,Atl	0.0006	0.0015	0.3817
Trans,Atl;Off,MI	0.0019	0.0018	1.056
Trans,MI;Off,Atl	0.0031	0.0012	2.467
Trans,Off;MI,Atl	0.0012	0.0014	0.840