TITLE: Inconsistencies in measures of population differentiation: Insights into FST and coalescent analysis

AUTHOR INFORMATION:

Gina Lamka

College of Forestry and Wildlife Sciences, Auburn University, Auburn, AL

[gfl0003@auburn.edu](mailto:gfl0003@auburn.edu)

KEYWORDS: coalescent, FST, gene\*, population, diverge

DECLARATIONS:

**Funding:** This work was supported by a Presidential Graduate Research Fellowship from Auburn University.

**Conflicts of interest:** The author declares no conflict of interest.

**Availability of data and material:** Supplemental material for this article may be found in the appendix.

**Code availability:** All processing and analysis steps executed on these data are stored in bash-scripts on the project GitHub page at https://github.com/ginalamka/Meta-Analysis

ABSTRACT:

As genetic exchange decreases in changing populations, genetic mutations may develop within subpopulations, leading to overall differences in population structure that can be strengthened by selection. Migration may still be evident in diverging populations, as identified through genetic techniques. Here, I present two variables that describe population divergence in existing literature, FST and coalescent-based migration rate, and compare the values generated for populations of wild mammals. Surprisingly, there are significant deviations in the way these two variables describe the same populations, with species and habitat contributing to the differences calculated here. The results presented identify discordant measures of population differentiation, which could lead to management decisions that do not reflect the true structure of these wild populations.

INTRODUCTION:

When one large population is subdivided by geographic isolation due to environmental or anthropogenic change, subpopulation structure can lead to differing allele frequencies. Populations diverge as a response to halted genetic exchange, occurring because ancestral species accumulate independent genetic mutations over time, resulting in reproductive isolation. If two or more subpopulations generate different allele frequencies, heterozygosity will decrease, causing subdivided populations to appear more inbred (Wahlund Effect; de Meeus 2018). As populations diverge and genetic mixing stops occurring over many generations, speciation may occur between the separated populations. Life history traits, abiotic and biotic environmental factors, selection, population size, and mutation rate are factors that contribute to overall speciation (Nei and Nozawa 2011; Rundle and Rowe 2018).

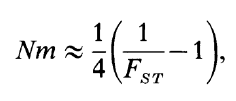
Inbreeding, speciation, and population divergence are commonly identified using genetic techniques, both from the macroevolutionary scale of diversification to the analysis of specific genomic regions for subpopulation structure evaluations (Castillo and Barbash 2017). Wright (1950) outlined a set of measures, *F* statistics, to examine how subdivided populations deviate from each other. Arguably the most useful for identifying the overall divergence among subpopulations is *FST* (Whitlock 2011). Similar measures such as RST, GST, QST, Jost’s D, and ΦST are utilized to empirically estimate genetic divergence of populations, though the most appropriate measure is still up for debate (Nei 1973; Michalakis and Excoffier 1996; Jost 2008, 2009; Kronholm et al. 2010; Bird et al. 2011; Meirmans and Hedrick 2011; Whitlock 2011). Similarly, statistic measures can be used to compare populations with a known ancestor to examine the maximum likelihood of allele merging with mutation rate, population size, and recombination rates in mind over many simulations (coalescent analysis; Beerli 1998; Neilsen and Wakeley 2001; Wang and Whitlock 2003; Cayuela et al. 2018). Coalescent approaches contain a multitude of variables to describe population divergence (migration rate m, Slatkin 1987; Tajima's D, Tajima 1989; Weir and Cockerham’s theta, Weir and Cockerham 1984; Nei's D, Nei 1978; Fu and Li's F, Fu and Li 1993; Cavalli-Sforza and Edwards’ Dc, Cavalli-Sforza and Edwards 1967; Watterson's theta, Watterson 1975). Due to the many diversity estimates currently available in the literature, managers must decide the best measure of population diversity estimates.

The differences among divergent populations can lead to differences in genetics, influencing morphological, behavioral, and physiological changes (Brumm et al. 2010). Perhaps the most classic example is the diverging populations of Darwin’s finches in the Galápagos (Farrington et al. 2014). As islands and other fragmented populations diversify, character trait differences such as beak shape and body size develop as an artifact of natural selection and random mutation (Grant and Grant 2008). Hoeck et al. (2010) utilized both microsatellite analysis and coalescent-based modeling to add to the growing knowledge of managing species in the Galápagos, though the results of both analyses are non-concordant, with microsatellite data suggesting high divergence, while coalescent approaches suggest the potential of genetic mixing.

The results presented here compare two genetic divergence estimators, FST and coalescent-based migration rate m, to determine if these measures – which are treated somewhat synonymously in the literature – describe population divergence in a similar way. To date, there have been no other studies that directly compare these estimates over the breadth of taxa included here. The overall null hypothesis of this study is that both direct and coalescent measures have similar results when describing diverging populations.

METHODS:

Web of Science was used to search the existing literature to compare FST with coalescent based approaches. The search string using the terms coalescent\*, FST, SNP, diverg\*, microsat\*, and genome produced 606 applicable studies (Appendix I). Literature was refined to only include articles and early access, while excluding proceedings and reviews. Studies were screened by excluding taxa that are not included in the Mammalia clade, leaving 168 potential studies. Following an examination of the full-text, many studies were excluded with reason, leaving a total of seven articles that fit the study criteria (Appendix III). Data collected from relevant literature included sample size, number of populations, genus and specific epithet, FST, and a coalescent divergence estimate (migration rate m). In the case of multiple migration rates or FST values for subpopulations, values were averaged to compare available population measures more directly. Authors, title, journal, and year published are noted. Additional variables included were allelic richness, observed and expected heterozygosity, nucleotide diversity, the country that samples were collected from, genetic marker type, and number of loci.

So that a direct comparison could be made among population divergence estimates, the quantitative FST was scaled and converted to the effective number of migrants (Nm), a measure more similar to the coalescent migration rate (Wright 1950).  (1)

Differences in the effective number of migrants and migration rate were calculated using a linear regression and random effects models using R (version 4.0.3) and Rstudio (version 1.4.1103). To meet normality assumptions, both variables were log transformed for each study. The log response ratio (lrr) was calculated using the equation:

(2)

where m is the migration rate collected from the literature and Nm is the effective number of migrants calculated from FST. The log response ratio was interpreted as the relationship between these two variables, with the expectation that if the ratio is zero, these two variables describe the populations similarly. However, if the ratio is nonzero, the results would suggest differences in the population divergence estimated from these two variables.

Several random effects models were examined, first without moderators, then with the number of loci and study as moderators in separate analyses using the *metafor* package (version 2.4-0). Because there was evidence suggesting that the differences in study were influenced by species, habitat (terrestrial versus marine) was set as a random factor in another moderator analysis. Following, to determine the strength of the relationship between m and Nm, a linear regression was conducted with Nm as the predictor and m as the response variable.

RESULTS:

A surprisingly low number of seven studies included both FST and migration rate values and were included in these analyses. From these, data were gathered for 26 population comparisons of seven species (Appendix II). The log response ratio was calculated as an unweighted effect size due to inconsistencies in reporting standard error and subpopulation sample size in the literature. The random effects model with lrr as the response variable detected significant differences, with ratios both above and below the zero mark (Fig. 1; estimate = -0.7572, z score = -3.1501, p = 0.0016). No significant effects were detected when using the number of microsatellite loci analyzed (Fig. 2, estimate = -0.0452, z-score = -1.0339, p = 0.3012). Differences in effect sizes were driven by the study the comparisons were taken from, which may be an artifact of species, as each study contained multiple population comparisons within the same species (Fig. 3; estimate = 0.0042, z-score = 3.1254, p = 0.0018). To determine what may be driving the differences within species, each species was classified as either a marine mammal or terrestrial mammal, with differences suggesting that habitat plays a role in altering these population divergence estimates (Fig. 4; estimate = 1.1143, z-score = 2.5534, p = 0.0107). As presented here, marine mammals had inconsistent reported values, while terrestrial FST and migration rate values were similar. A linear relationship among migration rate and effective number of migrants per generation was not evident (Fig. 5; adjusted r2 = -0.03071, p = 0.6181).

DISCUSSION:

Although the number of genetic markers can affect the quality of the genetic sample and increase uncertainty in population divergence estimates (Edwards and Beerli 2000; Kalinowski 2002; Sovic et al. 2014), the sample size here was sufficient to correctly identify population differentiation measures. Even though differentiation in the genome depends on locus-specific divergence (McKay and Latta 2002; Nosil et al. 2012), the microsatellite loci selected in the studies included here have not contributed to difference in population divergence measures. Microsatellites, neutral sections of non-coding DNA, are expected to be equally affected by migration and genetic divergence on all loci, and can be supported by coalescent theory (Latta 2003). Therefore, it is unsurprising that number of loci did not significantly affect the ratio of FST to migration rate.

Effect size differences between FST and migration rate were primarily driven by species, with habitat differences leading to the greatest separation in the data. Although habitat is a factor that can divide these species, migration distance may be an artifact of this difference, as marine mammals tend to migrate greater distances than the terrestrial-bound species that were present in this study (Carr et al. 2003; Luschi 2013; Joly et al. 2019). Marine species have been documented to exhibit high gene flow and little interpopulation genetic divergence relative to the high genetic diversity within the species, which can explain why the calculated effective number of migrants are greater than migration rate (Gyllensten 1985; Carr et al. 2003). Further, discrepancies between estimated effective and actual population numbers have been documented in marine animals, especially due to a bias in population sampling (Hedgecock 1994; Hare et al. 2011; Husemann et al. 2016; Nunney 2016; Marandel et al. 2019). Unfortunately, due to uncommon reporting of variances in these data, bias in this sample was not tested. Therefore, the discrepancies within FST and migration rate comparisons seem to be attributed to life history traits of these species rather than bias in population sampling, though further investigation would be necessary to draw a conclusion.

Charlesworth (1998) suggested that absolute measures of divergence are necessary when comparing different species, especially since FST is “inherently dependent on the extent of within-population diversity”. Null alleles are also commonly found in microsatellite genotyping, especially in nonequilibrium populations, which can lead to biased estimates of population differentiation (Chapuis and Estoup 2006; Van Oosterhout et al. 2006). This problem has been detected before, as genetic analysis of Antarctic krill indicated non-significant FST differentiation, while coalescent simulations detected significant divergence estimates (Bortolotto et al. 2011). When migration rates based on FST are confounded by coalescent values, offspring distribution may cause differential measures of genetic heterogeneity (Eldon and Wakeley 2009). These differences have been so evident, that Müller et al. (2017) suggest structuring coalescent simulations without inferring migration rates. Presently, the divergence of these two measures, FST and coalescent approaches, are still evident.

Perhaps the most surprising result here is that there was no relationship between FST and migration rate when compared via regression analysis. Here, I present data indicating that these two measures of population differentiation describe wild populations differently. Now, it is important for geneticists to further examine the differences in these measures and come to an agreement on the best measures of population divergence. Further studies should compare these results with those of other coalescent variables that currently exist in the literature. Additionally, further examination of these patterns should be applied to a greater breadth of taxa, perhaps with inspection into phylogenetic influences of these divergence differences.

REFERENCE LIST:

Beerli P (1998) Estimation of migration rates and population sizes in geographically structured populations. Adv Mol Ecol 306:39–53

Bird C, Karl S, Mouse P, Toonen R (2011) Detecting and measuring genetic differentiation. 31–55. https://doi.org/10.1201/b11113-4

Bortolotto E, Bucklin A, Mezzavilla M, et al (2011) Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill Euphausia superba

Brumm H, Farrington H, Petren K, Fessl B (2010) Evolutionary dead end in the Galápagos: Divergence of sexual signals in the rarest of Darwin’s finches. PLoS One 5:. https://doi.org/10.1371/journal.pone.0011191

Carr MH, Neigel JE, Estes JA, et al (2003) Comparing marine and terrestrial ecosystems: Implications for the design of coastal marine reserves. Ecol Appl 13:90–107. https://doi.org/10.1890/1051-0761(2003)013[0090:cmatei]2.0.co;2

Castillo DM, Barbash DA (2017) Moving Speciation Genetics Forward: Modern Techniques Build on Foundational Studies in Drosophila. https://doi.org/10.1534/genetics.116.187120

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 19:233–257. https://doi.org/10.2307/2406616

Cayuela H, Rougemont Q, Prunier JG, et al (2018) Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. Mol Ecol 27:3976–4010. https://doi.org/10.1111/mec.14848

Chapuis M-P, Estoup A (2006) Microsatellite Null Alleles and Estimation of Population Differentiation. Mol Biol Evol 24:621–631. https://doi.org/10.1093/molbev/msl191

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. Mol Biol Evol 15:538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

de Meeus T (2018) Revisiting FIS, FST, Wahlund effects, and null alleles. J Hered 109:446–456. https://doi.org/10.1093/jhered/esx106

Edwards S V, Beerli P (2000) PERSPECTIVE: GENE DIVERGENCE, POPULATION DIVERGENCE, AND THE VARIANCE INCOALESCENCE TIME IN PHYLOGEOGRAPHIC STUDIES. Evolution (N Y) 54:1839–1854

Eldon B, Wakeley J (2009) Coalescence times and FST under a skewed offspring distribution among individuals in a population. Genetics 181:615–629. https://doi.org/10.1534/genetics.108.094342

Farrington HL, Lawson LP, Clark CM, Petren K (2014) The evolutionary history of Darwin’s finches: Speciation, gene flow, and introgression in a fragmented landscape. Evolution (N Y) 68:2932–2944. https://doi.org/10.1111/evo.12484

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709. https://doi.org/10.1093/genetics/133.3.693

Grant PR, Grant BR. 2008. How and Why Species Multiply. The Radiation of Darwin’s Finches. Princeton NJ: Princeton University Press.

Gyllensten U (1985) The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous, and freshwater species. J Fish Biol 26:691–699. https://doi.org/10.1111/j.1095-8649.1985.tb04309.x

Hare MP, Nunney L, Schwartz MK, et al (2011) Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. Conserv Biol 25:438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: Genetics and evolution of aquatic organisms (Beaumont A, ed). 122–134.

Hoeck PEA, Beaumont MA, James KE, et al (2010) Saving darwin’s muse: Evolutionary genetics for the recovery of the floreana mockingbird. Biol Lett 6:212–215. https://doi.org/10.1098/rsbl.2009.0778

Husemann M, Zachos FE, Paxton RJ, Habel JC (2016) Effective population size in ecology and evolution. https://doi.org/10.1038/hdy.2016.75

Joly K, Gurarie E, Sorum MS, et al (2019) Longest terrestrial migrations and movements around the world. Sci Rep 9:15333. https://doi.org/10.1038/s41598-019-51884-5

Jost L (2008) GST and its relatives do not measure differentiation. Mol Ecol 17:4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Jost L (2009) D vs. GST: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Mol Ecol 18:2088–2091. https://doi.org/10.1111/j.1365-294X.2009.04186.x

Kalinowski S (2002) How many alleles per locus should be used to estimate genetic distances? Heredity (Edinb) 88:62–65. https://doi.org/10.1038/sj.hdy.6800009

Kronholm I, Loudet O, de Meaux J (2010) Influence of mutation rate on estimators of genetic differentiation - lessons from Arabidopsis thaliana. BMC Genet 11:7–9. https://doi.org/10.1186/1471-2156-11-33

Latta RG (2003) Gene flow, adaptive population divergence and comparative population structure across loci. New Phytol 161:51–58. https://doi.org/10.1046/j.1469-8137.2003.00920.x

Luschi P (2013) Long-Distance Animal Migrations in the Oceanic Environment: Orientation and Navigation Correlates. ISRN Zool 2013:631839. https://doi.org/10.1155/2013/631839

Marandel F, Lorance P, Berthelé O, et al (2019) Estimating effective population size of large marine populations, is it feasible? Fish Fish 20:189–198. https://doi.org/10.1111/faf.12338

McKay J, Latta R (2002) Genetic population divergence: markers and traits. Trends Ecol Evol 17:285–291. https://doi.org/10.1016/s0169-5347(02)02602-2

Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. Mol Ecol Resour 11:5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064. https://doi.org/10.1093/genetics/142.3.1061

Müller NF, Rasmussen DA, Stadler T (2017) The Structured Coalescent and Its Approximations. Mol Biol Evol 34:2970–2981. https://doi.org/10.1093/molbev/msx186

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations

Nei M, Nozawa M (2011) Roles of mutation and selection in speciation: From hugo de vries to the modern genomic era. Genome Biol Evol 3:812–829. https://doi.org/10.1093/gbe/evr028

Neilsen R, Wakeley J (2001) Distinguishing Migration From Isolation: AMarkov Chain Monte Carlo Approach. Genetics 158:885–896

Nosil P, Parchman TL, Feder JL, Gompert Z (2012) Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in Timema stick insects

Nunney L (2016) The effect of neighborhood size on effective population size in theory and in practice. Heredity (Edinb) 117:224–232. https://doi.org/10.1038/hdy.2016.76

Rundle HD, Rowe L (2018) The contribution of sexual selection to ecological and mutation-order speciation. Evolution (N Y) 72:2571–2575. https://doi.org/10.1111/evo.13599

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science (80- ) 236:787–792. https://doi.org/10.1126/science.3576198

Sovic MG, Kubatko LS, Fuerst PA, Paul Fuerst CA (2014) The effects of locus number, genetic divergence, and genotyping error on the utility of dominant markers for hybrid identification. Ecol Evol 4:462–473. https://doi.org/10.1002/ece3.833

Tajima F (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123:585–595

Van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Mol Ecol Notes 6:255–256. https://doi.org/10.1111/j.1471-8286.2005.01082.x

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163:429–446. https://doi.org/10.1093/genetics/163.1.429

Watterson GA (1975) On the Number of Segregating Sites in Genetical Models without Recombination. Theor Popul Biol 276:256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution (N Y) 38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Whitlock MC (2011) G′ST and D do not replace FST. Mol Ecol 20:1083–1091. https://doi.org/10.1111/j.1365-294X.2010.04996.x

Wright S (1950) The genetical structure of populations. Annals of Eugenics 15, 323-354

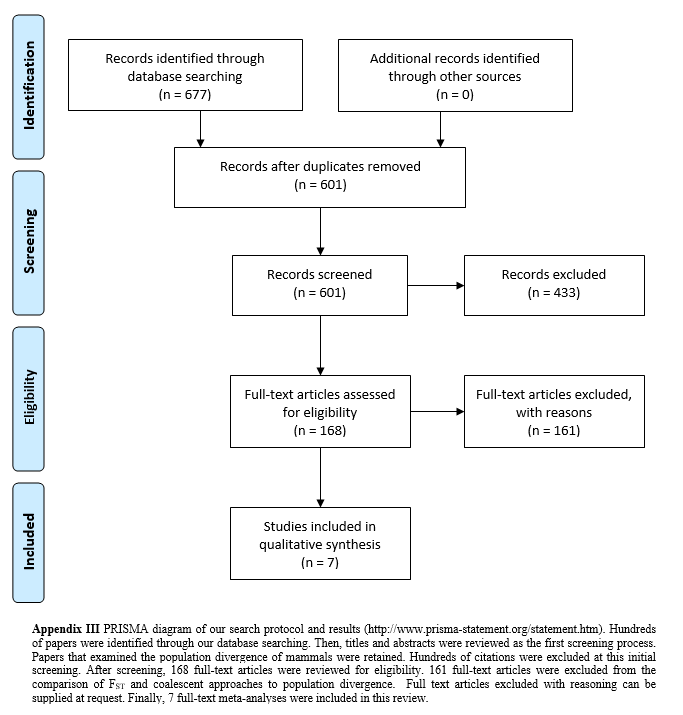
APPENDIX:

I. Keyword search string: (coalescent and FST) OR (coalescent\* and SNP and diverg\*) OR (coalescent and microsat\* and diverg\*) OR (coalescent and genome and diverg\*)

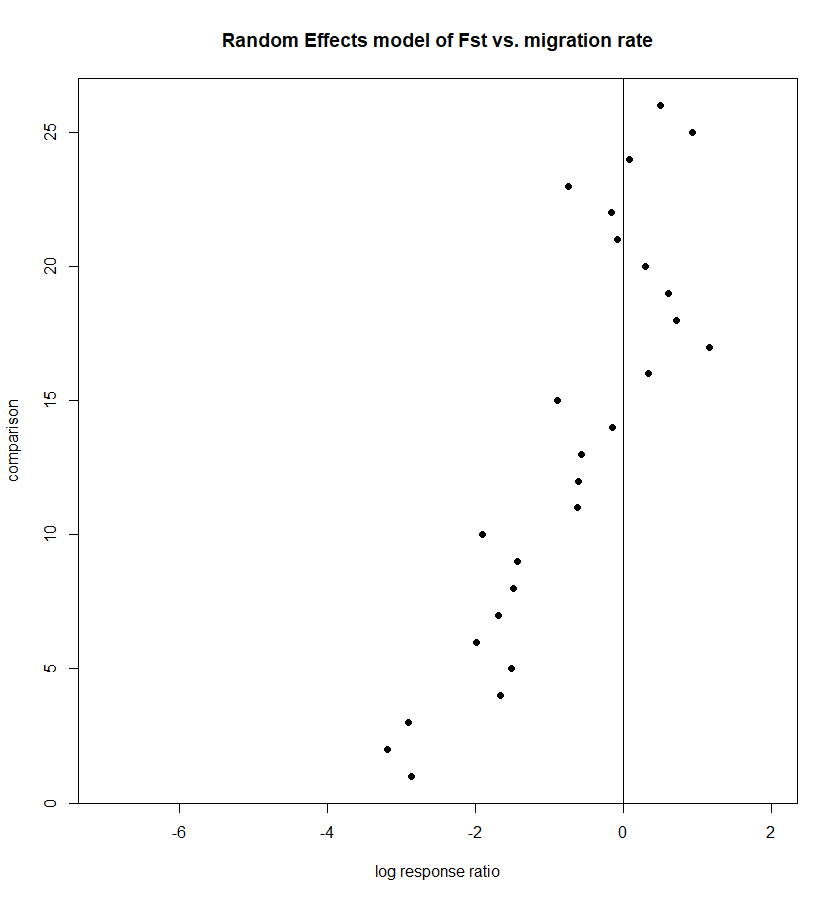
II. Table of relevant data

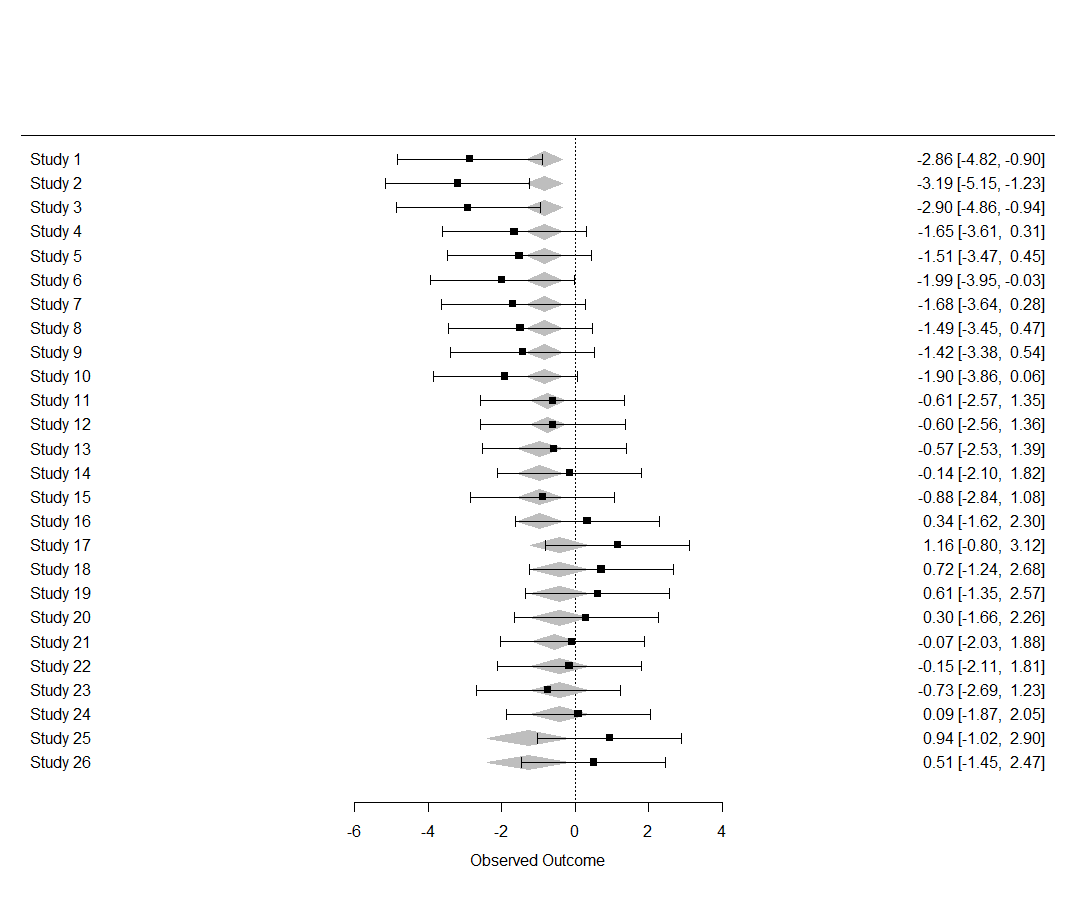


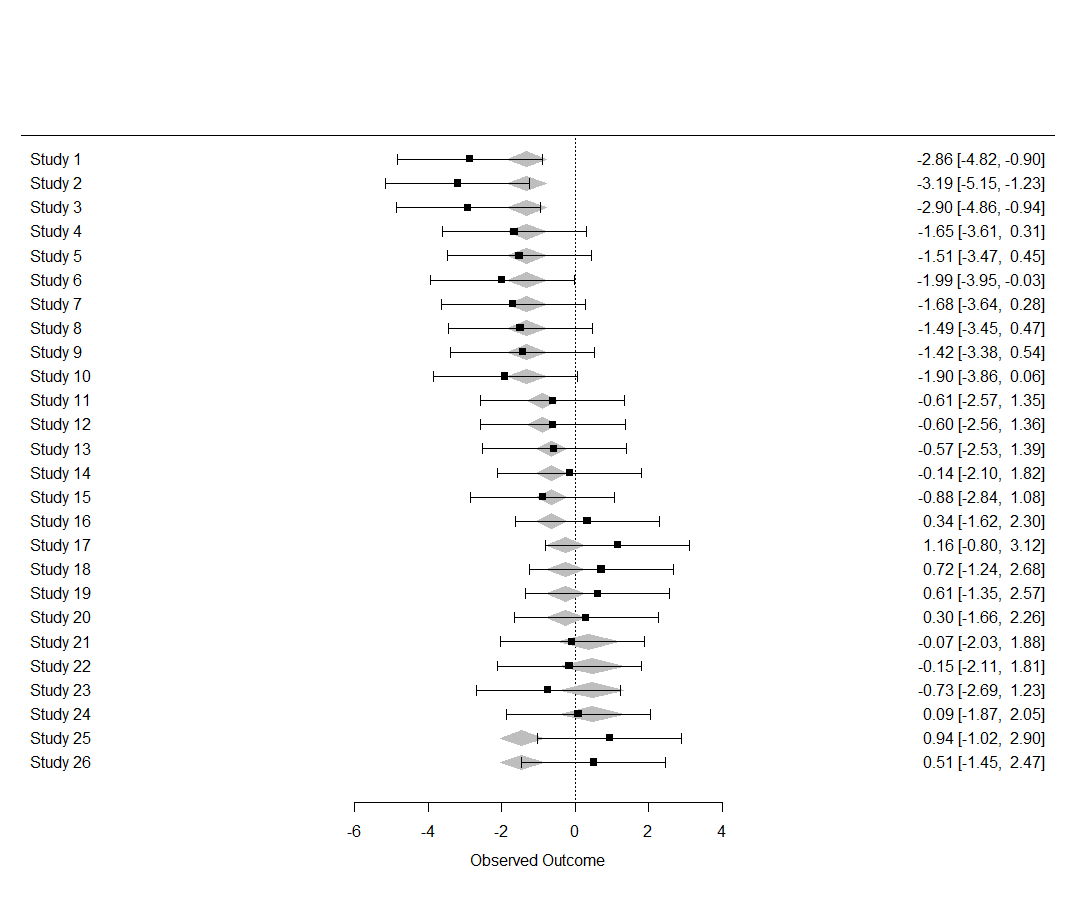
III. PRISMA flowchart

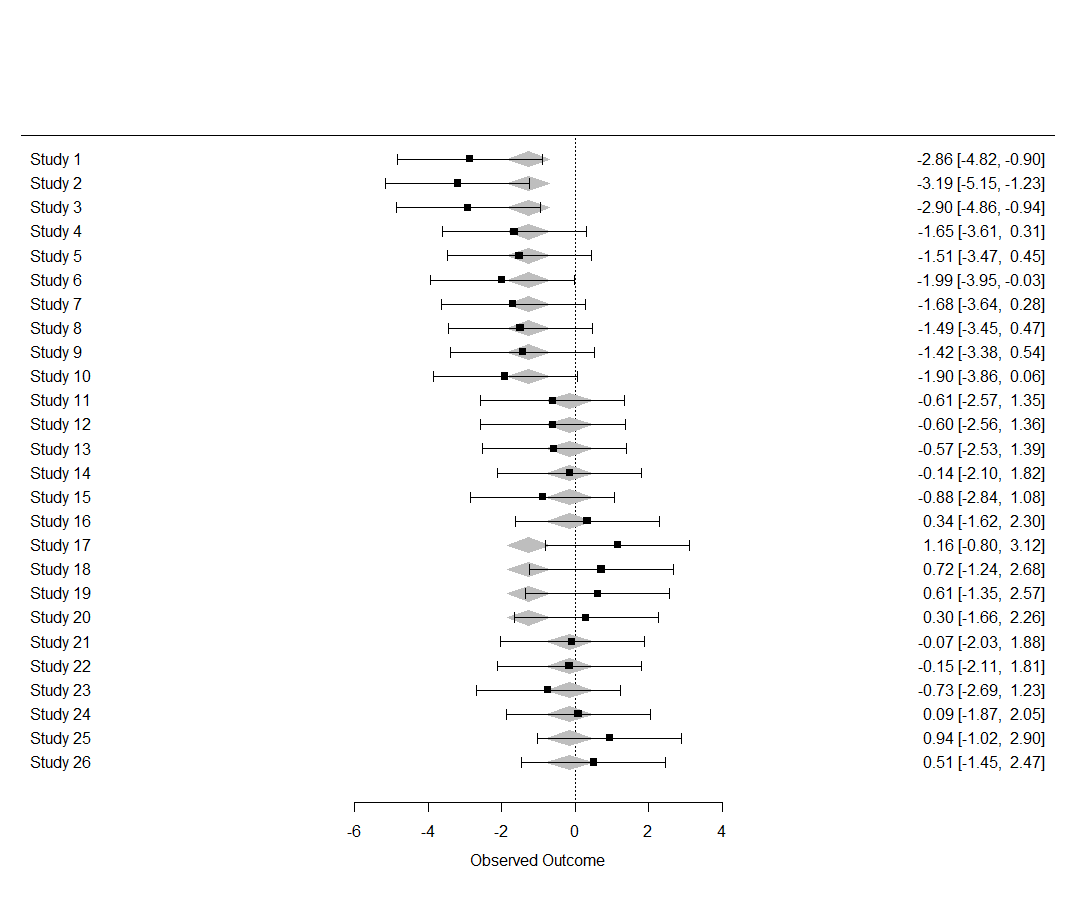


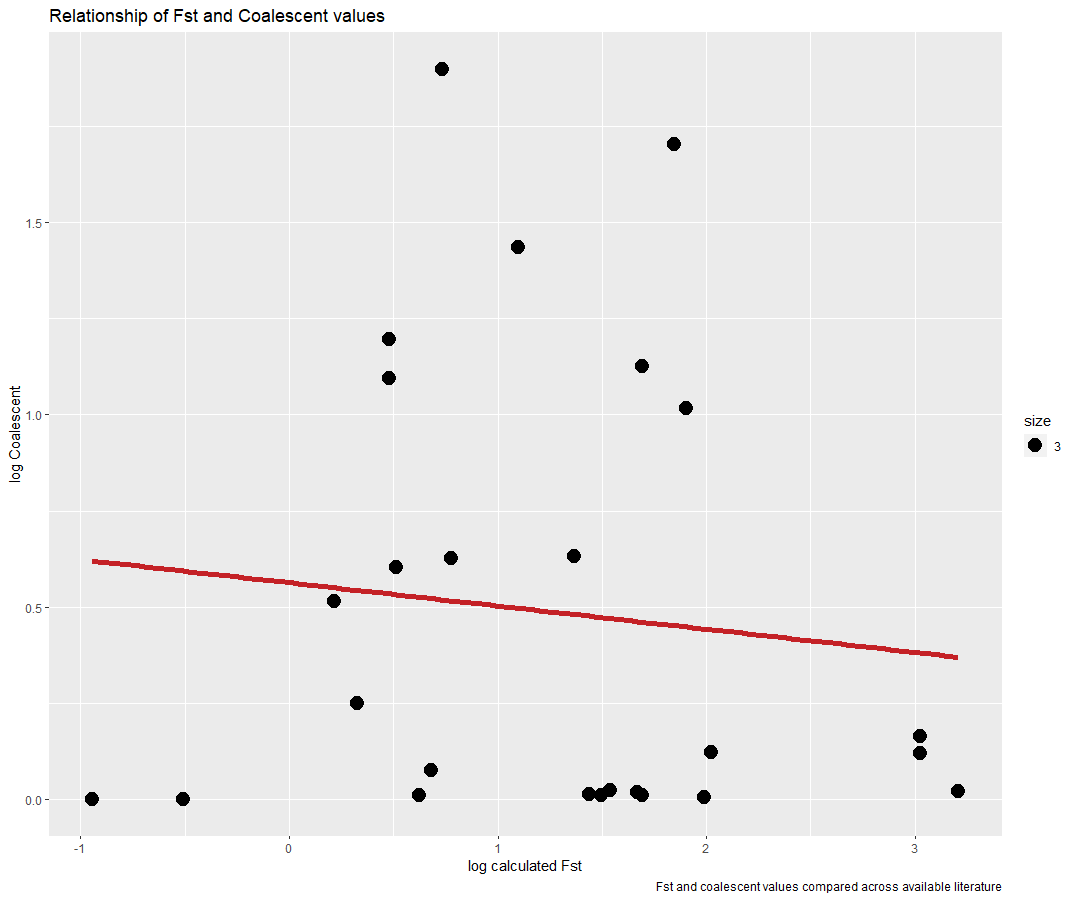
IV. Random effects models

  
Fig. 1. Random effects model with a log response ratio as the effect size. No variance estimates are reported. There is a negative overall response ratio that is significantly different from zero (p = 0.0016).

  
Fig. 2. Random effects model with number of loci used as a moderator. There are no significant differences from zero.

  
Fig. 3. Random effect model with study as a moderator. This could be an artifact of species, as each study contained one or more comparisons of migration from the same species. There is a significant positive response ratio (p = 0.0018).

  
Fig. 4. Random effects model with habitat as a moderator. There is a negative response ratio in marine species when compared to terrestrial species (p = 0.0107).

  
Fig. 5. Log linear regression between logged calculated FST and logged migration rate. There is no relationship between the variables (p = 0.618).