AlleleShift: An R package to predict and visualize population-level changes in allele frequencies in response to climate change

Roeland Kindt Corresp. 1

1 World Agroforestry, CIFOR-ICRAF, Nairobi, Kenya

Corresponding Author: Roeland Kindt
Email address: R.Kindt@CGIAR.org

Background. At any particular location, frequencies of alleles that are associated with adaptive traits are expected to change in future climates through local adaption and migration, including assisted migration (human-implemented when climate change is more rapid than natural migration rates). Making the assumption that the baseline frequencies of alleles across environmental gradients can act as a predictor of patterns in changed climates (typically future but possibly paleo-climates), a methodology is provided by *AlleleShift* of predicting changes in allele frequencies at the population level.

Methods. The prediction procedure involves a first calibration and prediction step through redundancy analysis (RDA), and a second calibration and prediction step through a generalized additive model (GAM) with a binomial family. As such, the procedure is fundamentally different to an alternative approach recently proposed to predict changes in allele frequencies from canonical correspondence analysis (CCA). The RDA step is based on the Euclidean distance that is also the typical distance used in Analysis of Molecular Variance (AMOVA). Because the RDA step or CCA approach sometimes predict negative allele frequencies, the GAM step ensures that allele frequencies are in the range of 0 to 1.

Results. *AlleleShift* provides data sets with predicted frequencies and several visualization methods to depict the predicted shifts in allele frequencies from baseline to changed climates. These visualizations include 'dot plot' graphics (function *shift.dot.ggplot*), pie diagrams (*shift.pie.ggplot*), moon diagrams (*shift.moon.ggplot*), 'waffle' diagrams (*shift.waffle.ggplot*) and smoothed surface diagrams of allele frequencies of baseline or future patterns in geographical space (*shift.surf.ggplot*). As these visualizations were generated through the *ggplot2* package, methods of generating animations for a climate change time series are straightforward, as shown in the documentation of *AlleleShift* and in the supplemental videos.

Availability. *AlleleShift* is available as an open-source R package from https://cran.r-project.org/package=AlleleShift and https://github.com/RoelandKindt/AlleleShift . Genetic input data is expected to be in the *adegenet::genpop* format, which can be generated from the *adegenet::genind* format. Climate data is available from various resources such as *WorldClim* and *Envirem*.

1 *AlleleShift:* **An R package to predict and visualize**
2 **population-level changes in allele frequencies in**
3 **response to climate change**
4
5 Roeland Kindt[1]

7 [1] World Agroforestry, CIFOR-ICRAF, Nairobi, Kenya
8
9 Corresponding Author:
10 Roeland Kindt[1]
11 UN Avenue, Gigiri, Nairobi, Kenya
12 Email address: R.Kindt@CGIAR.org
13
14 https://orcid.org/0000-0002-7672-0712

16
17

Abstract

21 **Background.** At any particular location, frequencies of alleles that are associated with adaptive
22 traits are expected to change in future climates through local adaption and migration, including
23 assisted migration (human-implemented when climate change is more rapid than natural
24 migration rates). Making the assumption that the baseline frequencies of alleles across
25 environmental gradients can act as a predictor of patterns in changed climates (typically future
26 but possibly paleo-climates), a methodology is provided by *AlleleShift* of predicting changes in
27 allele frequencies at the population level.

29 **Methods.** The prediction procedure involves a first calibration and prediction step through
30 redundancy analysis (RDA), and a second calibration and prediction step through a generalized
31 additive model (GAM) with a binomial family. As such, the procedure is fundamentally different
32 to an alternative approach recently proposed to predict changes in allele frequencies from
33 canonical correspondence analysis (CCA). The RDA step is based on the Euclidean distance that
34 is also the typical distance used in Analysis of Molecular Variance (AMOVA). Because the
35 RDA step or CCA approach sometimes predict negative allele frequencies, the GAM step
36 ensures that allele frequencies are in the range of 0 to 1.

37

38 **Results.** *AlleleShift* provides data sets with predicted frequencies and several visualization
39 methods to depict the predicted shifts in allele frequencies from baseline to changed climates.
40 These visualizations include 'dot plot' graphics (function *shift.dot.ggplot*), pie diagrams
41 (*shift.pie.ggplot*), moon diagrams (*shift.moon.ggplot*), 'waffle' diagrams (*shift.waffle.ggplot*) and
42 smoothed surface diagrams of allele frequencies of baseline or future patterns in geographical
43 space (*shift.surf.ggplot*). As these visualizations were generated through the *ggplot2* package,
44 methods of generating animations for a climate change time series are straightforward, as shown
45 in the documentation of *AlleleShift* and in the supplemental videos.

46

47 **Availability.** *AlleleShift* is available as an open-source R package from https://cran.r-
48 project.org/package=AlleleShift and https://github.com/RoelandKindt/AlleleShift . Genetic input
49 data is expected to be in the *adegenet::genpop* format, which can be generated from the
50 *adegenet::genind* format. Climate data is available from various resources such as *WorldClim*
51 and *Envirem*.

52

Introduction

53

54

55 There is clear evidence of anthropogenically induced climate change, with our planet facing a
56 climate emergency (Ripple *et al.*, 2020). Anticipating climate change, many countries are
57 developing National Adaptation Plans (NAPs; https://www4.unfccc.int/sites/napc). Specifically
58 for forests and trees, technical guidelines have recently been prepared on the integration of

59 forests, agroforestry and trees in the formulation and implementations of NAPs (Meybeck et al.,
60 2020). Stanturf et al. (2015) provide a practical framework and a stoplight tool to plan for
61 climate change mitigation and adaptation in forest and landscape restoration initiatives.
62
63 For loci involved in adaptation, shifts in allele frequencies (and changes in phenotypes as a
64 result) can be anticipated (Günther & Coop, 2013; Stange, Barrett & Hendry, 2020). Although
65 the methods involved are far from straightforward, statistical approaches such as Genome Wide
66 Association Studies (GWAS) and Environmental Association Analyses (EAA) can be applied to
67 genomic data to postulate genes and specific alleles involved in climate change responses
68 (Luikart et al., 2018; Anderson & Song, 2020; Waldvogel et al. 2020). *AlleleShift* predicts shifts
69 in allele frequencies for those loci predicted by GWAS or EAA to be associated with adaptive
70 traits.
71
72 The methodology used in *AlleleShift* exploits the analogy between analysing a community matrix
73 (consisting of sites as rows, species as columns and abundances as cell values) and a genetic
74 matrix (consisting of populations or individuals as rows, alleles as columns and allele counts as
75 cell values). As a consequence, the common constrained ordination method in the field of
76 community ecology of redundancy analysis (RDA; Legendre & Legendre 2012, pp. 629 - 661)
77 can be applied.
78
79 RDA is based on Euclidean distances and without explanatory variables is equivalent to principal
80 components analysis (PCA). Various recent studies of adaptative genetic variation have also
81 used the RDA methodology (e.g., Razgour et al.; 2019; Capblancq et al., 2020; Nelson,
82 Motamayor & Cornejo, 2020; Temunović et al., 2020). Analysis of a genetic matrix via
83 Euclidean distances is appropriate for several reasons:
84 •   Euclidean distances are also used in Analysis of Molecular Variance (AMOVA;
85   Excoffier, Smouse & Quattro, 1992; Meirmans & Liu, 2018; Michalakis & Excoffier,
86   1996) and it can be demonstrated (see examples for the `AlleleShift::amova.rda`
87   function and possibly also compare with AMOVA analysis in GenAlEx; Peakal &
88   Smouse, 2012) that RDA provides the same information on squared Euclidean distances
89   and mean squares as an AMOVA analysis.
90 •   In article S1, I demonstrate how Euclidean distances between `adegenet::genpop`
91   objects are linearly related to the Euclidean distances between the centroids obtained
92   from a PCA analysis of `adegenet::genind` objects. As a corollary, shifts of
93   populations can be understood as the average shift of individuals in ordination space.
94 •   Recently, I also showed (Kindt 2020a) how RDA can be directly interpreted in terms of
95   Sums-of-Squares of AMOVA by analysing distances from individuals to centroids and
96   among centroids.
97

98   The RDA model is calibrated with an explanatory data set that documents the environmental
99   values of populations in the baseline climate. The prediction function uses the values of
100 populations in the changed (future or past) climate as explanatory variables to predict changes in
101 allele counts and frequencies in the changed climate. As possibly negative allele counts can be
102 predicted by the calibrated RDA, as with the example data set shown below, a second calibration
103 step is implemented that guarantees that allele frequencies are within the biologically realistic
104 interval between 0 and 1.
105
106 By using RDA and avoiding negative allele frequencies, *AlleleShift* is fundamentally different
107 from the protocol developed by Blumstein et al. (2020) based on canonical correspondence
108 analysis (CCA; Legendre & Legendre, 20121, pp. 661 - 673). Besides the advantage of RDA to
109 reflect Euclidean distances, a disadvantage of CCA is that the interpretation of species scores in
110 ordination diagrams is more complex in showing the peak of its unimodal distribution against a
111 vector of an explanatory variable (see Fig. 3 in Ter Braak 1987; Fig. 11.9 in Legendre &
112 Legendre, 2012; or Figure 10.13 in Kindt & Coe, 2002).
113
114
115 **Materials & Methods**
116 **Data import**
117 Genetic response data (including a matrix with populations as rows and allele counts as
118 columns) for the calibration of the `AlleleShift::count.model` and prediction via
119 `AlleleShift::count.pred` is required to be in the `adegenet::genpop` format.
120 Individual-based data that are in the `adegenet::genind` format can be converted into the
121 genpop format via the `adegenet::genind2genpop` function. The *adegenet* and *poppr*
122 packages provide various methods of importing data from other software application formats into
123 the *genind* format, such as `adegenet::import2genind` and `poppr::read.genalex`.
124 Environmental data of populations, used as explanatory variables in redundancy analysis (RDA),
125 is expected to be provided as a `data.frame` with the same sequence of populations as the
126 genetic response data (this is a general requirement for community ecology methods in the *vegan*
127 and *BiodiversityR* packages; a check is available via
128 `BiodiversityR::check.datasets`). Whereas environmental data typically is baseline
129 and changed (bio)climatic data such as is available from WorldClim (Fick & Hijmans, 2017),
130 ENVIREM (Title & Bemmels, 2018), CHELSA (Karger et al., 2020) or PaleoClim (Brown et al.,
131 2018), it is also possible to expand input to other data available for species distribution modelling
132 (a recent overview of available data sets is provided by Booth, 2018).
133
134 **Data analysis**
135 Prior to calibrating the models that predict allele frequencies from bioclimatic explanatory data,
136 it is recommended to reduce the explanatory variables to a subset where the Variance Inflation
137 Factor (VIF) is below a predefined threshold for each variable. Such methods are also

138   recommended for regression analysis (Fox & Monette, 1992) and species distribution modelling
139   (Kindt, 2018). With this approach, it is easy to select the same variables from future data sets and
140   comparison with other studies may also become easier. VIF analysis, and an additional feature of
141   forcing the algorithm to keep preselected variables within the final subset, is available via
142   `AlleleShift::VIF.subset` (Table 1). This is a function that uses
143   `BiodiversityR::ensemble.VIF.dataframe` internally after a first step of removing
144   all explanatory variables that have correlations larger than the VIF threshold with the preselected
145   variables. There is an option to generate a correlation matrix for the final subset of variables
146   (Figure 1). I also recommend conducting the VIF analysis for the genetic response data (see
147   discussion and Figure 1). I further advise to remove any individuals with partially missing
148   genetic or (bio)climatic data prior to the analysis. As a sensitivity analysis, results could possibly
149   be compared with models calibrated with data where missing genotypes were replaced by
150   genotypes predicted by the `impute` function of the *LEA* package (Gain & François, 2021).
151
152   Prior to model calibration, I suggest checking (according to criteria described below) for the
153   shifts of populations in environmental space between the baseline and changed climates. This
154   can be achieved via function `AlleleShift::population.shift`, which draws arrows
155   between each population in the baseline and changed climate. There are alternatives of using
156   principal components analysis (PCA) or redundancy analysis (RDA) to generate the ordination
157   diagrams (Figure 2). What is important to check in the ordination graphs is whether populations
158   shift in a similar fashion, as that will facilitate the interpretation of predicted shifts in allele
159   frequencies. If some populations would show a different trend in the ordination graph, their allele
160   frequencies would also be expected to change in a different way.
161
162   With the selected genetic and explanatory data, model calibration can be done. In a first step, a
163   RDA model is fitted (`AlleleShift::count.model`) that can predict counts of alleles in
164   baseline or changed climates (`AlleleShift::count.pred`). The user has the option also to
165   obtain results for the canonical correspondence analysis procedure of Blumstein et al. (2020)
166   with the count model via argument `cca.model=TRUE`. In the second step, the predicted allele
167   counts for the baseline climate serve as explanatory variables for the calibration of a generalized
168   linear model (GAM via `mgcv::gam`; Wood, 2004) with the baseline allele frequencies as
169   response and a binomial family function (`AlleleShift::freq.model`). This procedure
170   ensures that predictions are within the realistic interval for frequencies between 0 to 1. Function
171   `AlleleShift::freq.pred` allows the prediction of allele frequencies for baseline and
172   changed climates. A simulation study (article S2, using a subset of a simulation study available
173   from Frichot & François, 2015) and the results below show that the predicted frequencies
174   approximate the expected frequencies well.
175
176   The output of the two steps (RDA followed by GAM) is a data.frame as shown in Table 2 (to fit
177   printing space available here, the data is shown in a transposed format where rows and columns

178   were swapped). All figures shown in this manuscript were obtained with the example data sets of
179   `AlleleShift::Poptri.genind` (individually-based allele counts),
180   `AlleleShift::Poptri.baseline.env` (climatic descriptors of the populations in the
181   baseline climate), `AlleleShift::Poptri.future.env` (climatic descriptors of the
182   populations in the future climate) and `AlleleShift::Poptri.loc` (geographical
183   coordinates of the populations). These data sets were converted from the example data sets
184   provided by Blumstein et al. (2020) for *Populus trichocarpa*. It can be seen for population
185   Nisqually in our case study that negative allele counts and frequencies are predicted for one of
186   the minor alleles in the RDA prediction step, but that the GAM step predicts the biologically
187   acceptable frequency of 0.027. Function `AlleleShift::freq.ggplot` (Figure 3) enables
188   a visual inspection of the power of the models to predict allele frequencies for the calibration
189   data.
190
191   **Visualizations**
192   *AlleleShift* generates various types of *ggplot2* (Wickham, 2016; version 3.3.2) graphics from the
193   output of `AlleleShift::freq.pred`. These include dot (Figure 4), pie or donut (Figure
194   5), moon (Figure 6) and waffle (Figure 7) graphics, and smoothed regression surfaces (Figure 8).
195   As an intermediate step to generate various of these graphics, helper functions such as
196   `waffle.baker` for `shift.waffle.ggplot` or `moon.waxer` for
197   `shift.moon.ggplot` prepare data for the main graphing function (Table 1). With default
198   settings, visualizations depict changes in allele frequencies for each allele in different panels,
199   internally via `ggplot2::facet_grid`. Setting argument `mean.change` to `TRUE`, the
200   graphics depict median or mean changes in allele frequencies.
201
202   Function `shift.surf.ggplot` plots populations in geographical space via their geographical
203   coordinates (longitude and latitude in Figure 8). The function fits a smoothed regression surface
204   for allele frequencies via `vegan::ordisurf`. This output is then further processed internally
205   within the function via `BiodiversityR::ordisurfgrid.long` (an overview of
206   generating *ggplot2* ordination diagrams via *vegan* and *BiodiversityR* is given in Kindt, 2020b;
207   these guidelines can be used to customize ordination graphs as shown in Figures 2 and 8).
208   Various options of fitting smoothed regression surfaces are available by providing additional
209   arguments to `shift.surf.ggplot`, such as the various `mgcv::smooth.terms` options
210   of thin plates, Duchon splines, cubic regression splines, P-splines, Markov Random Fields,
211   Gaussian process smooths, soap film smooths, splines on the sphere and random effects.
212
213   As graphics are generated with *ggplot2*, it is relatively easy to generate animated versions of
214   visualizations with *gganimate* (Pedersen & Robinson, 2020; version 1.0.7). Scripts for
215   generating animated versions for dot graphics, pie graphics and smoothed surfaces are provided
216   with the documentation of the respective functions. Supplemental MP4 videos (S3 – S5) are for a
217   time series that interpolates bioclimatic data between baseline and future climate in steps of five

218   years; these also extrapolate data far into the future (up to a million years, see the Results and
219   Discussion for my reason to do this). Other than time series, animated visualizations could also
220   be generated for different global circulation models (GCMs) or scenarios such as various shared
221   socio-economic pathways (SSPs) developed in the context of the sixth assessment report of the
222   Intergovernmental Panel on Climate Change .
223
224   **Results and Discussion**
225
226   *AlleleShift* predicts shifts in allele frequencies via RDA and GAM, an alternative pathway that
227   maintains Euclidean distances among populations and individuals (Article S1). It also avoids
228   making negative frequency predictions as what may occur with the protocol recently devised by
229   Blumstein et al. (2020). My methodology however faces the same limitations of data
230   requirements as discussed by Blumstein et al. (2020) for their protocol in terms of the initial
231   identification of responsive markers. Key assumptions of the Blumstein et al. (2020) protocol
232   apply also to my approach, including that allelic effects are independent and additive with no
233   epistasis or dominance (see also the discussion on epistasis, structural genomic variation and
234   epigenetics by Stange, Barrett & Hendry, 2020).
235
236   I recommend reducing both the explanatory variables and the response variables to subsets of
237   data with a maximum Variance Inflation Factor (VIF) of 20 or less (Ter Braak & Smilauer,
238   2002), for alleles to obtain better estimates of changes in their frequencies. For the allele counts
239   as response variables, if VIF would be larger than 20, I would use function `VIF.subset` in an
240   iterative procedure whereby earlier selected subsets of alleles are excluded, to generate different
241   subsets of alleles (but keeping variables for both A and B allele counts in each final subset). For
242   example, with 20 alleles X01A, X01B, X02A, X02B, … , X10A, X10B, the two subsets could be
243   {X01A, X01B, …, X05A, X05B} and {X06A, X06B, … , X10A, X10B}. With the various
244   subsets, shifts in allele frequencies can then be predicted, and finally predictions with all subsets
245   can be combined.
246
247   When predictions are made by *AlleleShift* into the future, and especially into novel climatic
248   conditions, it is warranted to consider the transferability of the calibrated models and ideally to
249   provide 'transferability metrics' that quantify prediction uncertainty (Yates et al., 2018). The
250   function of `environmental.novel` identifies populations that have novel climatic
251   conditions (values outside the baseline range) for at least one of the explanatory variables for the
252   changed climate. This function also calculates the probabilities of encountering the novel
253   conditions from the mean and standard deviation of the baseline ranges and then returns the
254   lowest of these probabilities as an indicator of the degree of novelty of the environmental
255   conditions.
256

257   For the animated graphics, as an extreme example of the caveats of the methodology, I made
258   projections one million years into the future. These projections followed earlier trends of the 21st
259   century and resulted in allele frequencies becoming fixed at 0 or 1, which is a biologically
260   possible scenario. At the same time, however, these simulations clearly illustrated that the
261   methodology through RDA is correlative as in correlative/phenomenological species distribution
262   models. In fact, the linear extrapolation of climate variables resulted in an environmental data set
263   where the mean annual temperature was above 30,000 degrees, which obviously is a biologically
264   irrelevant scenario. Thus, the method presented here should be used cautiously in novel climates
265   especially as predictions will be available (the model will not crash or give an error warning),
266   and should thus be used even more cautiously where differences between baseline and future
267   climates are large.
268
269   As an approach to obtain a better handle on the transferability of *AlleleShift* predictions into
270   future climates, I recommend to also estimate the habitat suitability via species distribution
271   models (SDMs). Where habitat suitability is below a certain threshold, for instance a threshold
272   where the sum of sensitivity and specificity is maximized (Liu, White & Newell, 2013), results
273   of predictions of *AlleleShift* could be annotated of having lower transferability. Similarly, where
274   evaluation strips as proposed by Elith et al. (2005) suggest that it is less likely that the habitat is
275   suitable within the changed range of environmental conditions, additional information could be
276   provided on a lower transferability. Well-documented methods of utilizing SDMs to predict
277   shifts in species habitat suitability are available in the literature, including recent examples that
278   use the ensemble suitability modelling framework available in *BiodiversityR* (e.g., de Sousa et
279   al., 2019; Fremout et al., 2020; Kindt, 2018; Ranjitkar et al., 2014). For organisms such as trees,
280   correlative SDMs remain the best available method of predicting future species suitability,
281   whereas the limitations of these methods may not be as great as has been suggested (Booth,
282   2016; Booth, 2018). What is also attractive about SDMs is that a wider set of presence
283   observations are likely to be available than those populations that have been studied genetically.
284   Presence data are available from open-source databases such as GBIF or the Botanical
285   Information and Ecology Network (Enquist et al., 2016). Further to the collation of a larger set of
286   presence observations, application of SDMs should be straightforward using the same
287   (bio)climatic data sets as applied in *AlleleShift*. The approach of combining the results of
288   *AlleleShift* with SDM is somewhat similar to the method applied by Aguirre-Liguori et al. (2019)
289   to develop species distribution models for alleles. In my proposal, however, the predictions of
290   allele frequencies and SDM are done independently, and ideally with an expanded point presence
291   data set for SDM.
292
293   A straightforward and practical expansion of the methodology I have proposed is to tree seed
294   sourcing programmes (Broadhurst et al., 2008), possibly for developing schemes of human-
295   assisted geneflow *sensu* Aitken & Whitlock (2013). This is important for ensuring the matching
296   of planting materials to the conditions of planting sites (Cernansky, 2018; Roshetko et al., 2018;

297 Kettle et al., 2020). For specific planting sites and planting times (considering the perennial
298 nature of trees, climate change during the production cycle should be considered) of interest, the
299 prediction methods can readily provide the predicted allele frequencies needed for adaptation.
300 Theoretically, based on the similarity between predicted allele frequencies and those of available
301 source populations, the best matching source can then be selected. Similar approaches to devise
302 transfer and conservation schemes in the face of climate change can be employed for other
303 organisms than plants (Fitzpatrick & Edelsparre, 2018; Rochat et al., 2021).
304

Conclusions

306 The R package *AlleleShift* provides a set of functions that allow the prediction of allele
307 frequencies from baseline, future and past (bio)climatic explanatory variables via redundancy
308 analysis (RDA) and generalized additive models (GAM). Various visualizations are provided via
309 *ggplot2* and its extension packages such as *ggforce* and *gganimate*. At the time of submission of
310 this manuscript, no package was available for this set of tools. As with any other methodology
311 that projects into the future, it is important to reflect on transferability to novel climates.
312

Acknowledgements

314 The author thanks Ian Dawson (CIFOR-ICRAF) for reviewing the article prior to submission and
315 three anonymous reviewers who provided excellent suggestions to improve the manuscript. He
316 also thanks additional colleagues from CIFOR-ICRAF and the University of Copenhagen for
317 useful discussions on the applications of this package, including Lars Graudal, Prasad Hendre,
318 Ramni Jamnadass and Jens-Peter B. Lillesø. The CGIAR Research Program on Forests, Trees,
319 and Agroforestry (supported by the CGIAR Trust Fund) and the Provision of Adequate Tree
320 Seed Portfolios project (supported by Norway's International Climate and Forest Initiative
321 through the Royal Norwegian Embassy in Ethiopia) supported the author's time on this project.

322

References

324 Aguirre-Liguori, J.A., Ramírez-Barahona, S., Tiffin, P. and Eguiarte, L.E., 2019. Climate change
325   is predicted to disrupt patterns of local adaptation in wild and cultivated maize.
326   Proceedings of the Royal Society B, 286(1906), p.20190486.
327   https://doi.org/10.1098/rspb.2019.0486
328 Aitken, S.N., Whitlock, M.C. (2013). Assisted gene flow to facilitate local adaptation to climate
329   change. Annual review of ecology, evolution, and systematics, 44.
330   https://doi.org/10.1146/annurev-ecolsys-110512-135747
331 Anderson, J.T. and Song, B.-H. (2020). Plant adaptation to climate change—Where are we?. J.
332   Syst. Evol., 58: 533-545. https://doi.org/10.1111/jse.12649
333 Blumstein et al. 2020. Protocol for Projecting Allele Frequency Change under Future Climate
334   Change at Adaptive-Associated Loci. https://doi.org/10.1016/j.xpro.2020.100061

335 Booth, T.H. (2016). Estimating potential range and hence climatic adaptability in selected tree
336   species. Forest Ecology and Management, 366, pp.175-183.
337   https://doi.org/10.1016/j.foreco.2016.02.009
338 Booth, T.H. (2018). Species distribution modelling tools and databases to assist managing forests
339   under climate change. Forest Ecology and Management 430: 196-203. URL
340   https://doi.org/10.1016/j.foreco.2018.08.019 .
341 Bramson, M. (2019). gggibbous: Moon Charts, a Pie Chart Alternative. R package version 0.1.0.
342   https://CRAN.R-project.org/package=gggibbous
343 Broadhurst, L.M., Lowe, A., Coates, D.J., Cunningham, S.A., McDonald, M., Vesk, P.A. and
344   Yates, C. (2008), Seed supply for broadscale restoration: maximizing evolutionary
345   potential. Evolutionary Applications, 1: 587-597. https://doi.org/10.1111/j.1752-
346   4571.2008.00045.x
347 Brown, J., Hill, D., Dolan, A. et al. (2018). PaleoClim, high spatial resolution paleoclimate
348   surfaces for global land areas. Sci Data 5, 180254. https://doi.org/10.1038/sdata.2018.254
349 Capblancq, T, Morin, X, Gueguen, M, Renaud, J, Lobreaux, S, Bazin, E. Climate-associated
350   genetic variation in Fagus sylvatica and potential responses to climate change in the
351   French Alps. J Evol Biol. 2020; 33: 783– 796. https://doi.org/10.1111/jeb.13610
352 Cernansky, R. (2018). How to plant a trillion trees. Nature 560: 542-544.
353   https://doi.org/10.1038/d41586-018-06031-x
354 de Sousa, K., van Zonneveld, M., Holmgren, M. et al. (2019). The future of coffee and cocoa
355   agroforestry in a warmer Mesoamerica. Sci Rep 9, 8828. https://doi.org/10.1038/s41598-
356   019-45491-7
357 Elith, J., Ferrier, S., Huettmann, F. and Leathwick, J., 2005. The evaluation strip: a new and
358   robust method for plotting predicted responses from species distribution models.
359   Ecological modelling, 186(3), pp.280-289.
360   https://doi.org/10.1016/j.ecolmodel.2004.12.007
361 Enquist, B.J., Condit, R., Peet, R.K., Schildhauer, M. and Thiers, B.M., 2016.
362   Cyberinfrastructure for an integrated botanical information network to investigate the
363   ecological impacts of global climate change on plant biodiversity (No. e2615v2). PeerJ
364   Preprints. https://doi.org/10.7287/peerj.preprints.2615v2
365 Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992). Analysis of molecular variance inferred
366   from metric distances among DNA haplotypes: application to human mitochondrial DNA
367   restriction data. Genetics, 131, 479-491.
368 Fick, S.E., Hijmans, R.J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for
369   global land areas. International Journal of Climatology 37 (12): 4302-4315.
370   https://www.worldclim.org
371 Fitzpatrick, M.J. and Edelsparre, A.H., 2018. The genomics of climate change. Science,
372   359(6371), pp.29-30. https://science.sciencemag.org/content/359/6371/29.full
373 Fox, J., Monette, G., 1992. Generalized collinearity diagnostics. J. Am. Stat. Assoc. 87, 178e183.
374   https://doi.org/10.1080/01621459.1992.10475190 .

375 Fremout, T, Thomas, E, Gaisberger, H, et al. (2020). Mapping tree species vulnerability to
376   multiple threats as a guide to restoration and conservation of tropical dry forests. Glob
377   Change Biol.; 26: 3552– 3568. https://doi.org/10.1111/gcb.15028
378 Frichot, E. and François, O. (2015), LEA: An R package for landscape and ecological
379   association studies. Methods Ecol Evol, 6: 925-929. https://doi.org/10.1111/2041-
380   210X.12382
381 Gain, C. and François, O. (2021), LEA 3: Factor models in population genetics and ecological
382   genomics with R. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.13366 .
383 Günther, T. and Coop, G., 2013. Robust identification of local adaptation from allele
384   frequencies. Genetics, 195(1), pp.205-220.
385   https://dx.doi.org/10.1534%2Fgenetics.113.152462
386 Kamvar ZN, Tabima JF, Grünwald NJ. (2014) Poppr: an R package for genetic analysis of
387   populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.
388   https://doi.org/10.7717/peerj.281
389 Karger, D.N., Schmatz, D.R., Dettling, G. et al. High-resolution monthly precipitation and
390   temperature time series from 2006 to 2100. Sci Data 7, 248 (2020).
391   https://doi.org/10.1038/s41597-020-00587-y
392 Kettle, C.J., Atkinson, R., Boshier, D., Ducci, F., Dawson, I., Ekué, M., Elias, M., Graudal, L.,
393   Jalonen, R., Koskela, J. and Monteverdi, M.C., 2020. Priorities, challenges and
394   opportunities for supplying tree genetic resources. Restoring the Earth-The next decade:
395   Unasylva No. 252-Vol. 71 2020/1, 252(1), p.51.
396   http://www.fao.org/3/cb1600en/CB1600EN.pdf
397 Kindt, R. (2018). Ensemble species distribution modelling with transformed suitability values.
398   Environmental Modelling & Software 100: 136-145.
399   https://doi.org/10.1016/j.envsoft.2017.11.009
400 Kindt, R. (2020a). Ordination graphs with vegan, BiodiversityR and ggplot2.
401   https://rpubs.com/Roeland-KINDT
402 Kindt, R. (2020b). Analysis of Molecular Variance (AMOVA) with vegan and BiodiversityR,
403   including a graphical method to identify potential migrants. https://rpubs.com/Roeland-
404   KINDT
405 Kindt, R., Coe, R. (2005). Tree Diversity Analysis. A manual and software for common
406   statistical methods for ecological and biodiversity studies. https://CRAN.R-
407   project.org/package=BiodiversityR
408 Lasky, J.R., Forester, B.R. and Reimherr, M., 2018. Coherent synthesis of genomic associations
409   with phenotypes and home environments. Molecular Ecology Resources, 18(1), pp.91-
410   106. https://doi.org/10.1111/1755-0998.12714
411 Legendre, P., Legendre, L., 2012. Numerical ecology. Elsevier.
412 Liu, C., White, M. and Newell, G. (2013), Selecting thresholds for the prediction of species
413   occurrence with presence-only data. J. Biogeogr., 40: 778-789.
414   https://doi.org/10.1111/jbi.12058

415   Luikart G., Kardos M., Hand B.K., Rajora O.P., Aitken S.N., Hohenlohe P.A. (2018) Population
416   Genomics: Advancing Understanding of Nature. In: Rajora O. (eds) Population
417   Genomics. Population Genomics. Springer, Cham.
418   https://doi.org/10.1007/13836_2018_60

419   Meirmans, P., Liu, S. (2018). Analysis of Molecular Variance (AMOVA) for Autopolyploids
420   Front. Ecol. Evol., 23. https://doi.org/10.3389/fevo.2018.00066

421   Meybeck, A., Gitz, V., Wolf, J. and Wong, T. 2020. Addressing forestry and agroforestry in
422   National Adaptation Plans – Supplementary guidelines. Place of publication,
423   Bogor/Rome. FAO and FTA. https://doi.org/10.4060/cb1203en

424   Michalakis, Y., Excoffier, L. (1996). A generic estimation of population subdivision using
425   distances between alleles with special reference for microsatellite loci. Genetics 142,
426   1061–1064.

427   Nelson, J.T., Motamayor, J.C. and Cornejo, O.E., 2020. Environment and pathogens shape local
428   and regional adaptations to climate change in the chocolate tree, *Theobroma cacao* L.
429   Molecular Ecology. https://doi.org/10.1111/mec.15754

430   Peakall R and Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic
431   software for teaching and research - an update.
432   https://doi.org/10.1093/bioinformatics/bts460.

433   Pedersen, T.L. (2020). ggforce: Accelerating 'ggplot2'. R package version 0.3.2.
434   https://CRAN.R-project.org/package=ggforce

435   Pedersen, T.L., Robinson, D. (2020). gganimate: A Grammar of Animated Graphics. R package
436   version 1.0.7. https://CRAN.R-project.org/package=gganimate

437   Ranjitkar, S., Xu, J., Shrestha, K.K. and Kindt, R. (2014). Ensemble forecast of climate
438   suitability for the Trans-Himalayan Nyctaginaceae species. Ecological Modelling, 282: 8-
439   24. https://doi.org/10.1016/j.ecolmodel.2014.03.003

440   Razgour, O., Forester, B., Taggart, J.B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S.J.,
441   Novella-Fernandez, R., Alberdi, A. and Manel, S. (2019). Considering adaptive genetic
442   variation in climate change vulnerability assessment reduces species range loss
443   projections. Proceedings of the National Academy of Sciences, 116(21), 10418-10423.
444   https://doi.org/10.1073/pnas.1820663116

445   Rengefors, K., Gollnisch, R., Sassenhagen, I., Härnström Aloisi, K., Svensson, M., Lebret, K.,
446   Čertnerová, D., Cresko, W.A., Bassham, S. and Ahrén, D. (2021). Genome-wide SNP
447   markers reveal population structure and dispersal direction of an expanding nuisance
448   algal bloom species. Molecular Ecology. https://doi.org/10.1111/mec.15787

449   Ripple, W.J., Christopher Wolf, Thomas M Newsome, Phoebe Barnard, William R Moomaw, et
450   al. (2020). World Scientists' Warning of a Climate Emergency, BioScience 70: 8–12,
451   https://doi.org/10.1093/biosci/biz088

452   Rochat, E, Selmoni, O, Joost, S. (2021). Spatial areas of genotype probability: Predicting the
453   spatial distribution of adaptive genetic variants under future climatic conditions. Divers
454   Distrib. 00: 1– 15. https://doi.org/10.1111/ddi.13256

455 Roshetko, J.M., Dawson, I.K., Urquiola, J., Lasco, R.D., Leimona, B., Weber, J.C., Bozzano, M.,
456   Lillesø, J.P.B., Graudal, L. and Jamnadass, R. (2018). To what extent are genetic
457   resources considered in environmental service provision? A case study based on trees and
458   carbon sequestration. Climate and Development, 10(8), pp.755-768.
459   https://doi.org/10.1080/17565529.2017.1334620
460 Stange, M., Barrett, R.D. and Hendry, A.P. (2020). The importance of genomic variation for
461   biodiversity, ecosystems and people. Nature Reviews Genetics, pp.1-17.
462 Stanturf, J.A, Kant, P., Lillesø, J.-P.B., Mansourian, S., Kleine, M., Lars Graudal, L., Palle
463   Madsen, P. (2015). Forest Landscape Restoration as a Key Component of Climate
464   Change Mitigation and Adaptation. IUFRO World Series Volume 34. Vienna 72 p.
465 Temunović, M., Garnier-Géré, P., Morić, M., Franjić, J., Ivanković, M., Bogdan, S. and Hampe,
466   A. (2020). Candidate gene SNP variation in floodplain populations of pedunculate oak
467   (*Quercus robur* L.) near the species' southern range margin: Weak differentiation yet
468   distinct associations with water availability. Molecular Ecology, 29(13), pp.2359-2378.
469   https://doi.org/10.1111/mec.15492
470 Ter Braak, C.J. and Smilauer, P., 2002. CANOCO reference manual and CanoDraw for
471   Windows user's guide: software for canonical community ordination (version 4.5).
472 Ter Braak, C.J., 1987. The analysis of vegetation-environment relationships by canonical
473   correspondence analysis. Vegetatio, 69(1), pp.69-77.
474 Title P.O., Bemmels J.B. (2018). ENVIREM: an expanded set of bioclimatic and topographic
475   variables increases flexibility and improves performance of ecological niche modeling.
476   Ecography. 41:291–307. https://envirem.github.io/
477 Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R.,
478   Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O., Bergland, A., Flatt, T.,
479   Guillaume, F. and Pfenninger, M. (2020), Evolutionary genomics can improve prediction
480   of species' responses to climate change. Evolution Letters, 4: 4-18.
481   https://doi.org/10.1002/evl3.154
482 Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,
483   2016. https://ggplot2.tidyverse.org
484 Wood, S.N. (2004) Stable and Efficient Multiple Smoothing Parameter Estimation for
485   Generalized Additive Models, Journal of the American Statistical Association, 99: 467,
486   673-686. https://doi.org/10.1198/016214504000000980
487 Yates, K.L., Phil J. Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding,
488   A.H., Bamford, A.J., Ban, S., et al. (2018). Outstanding Challenges in the Transferability
489   of Ecological Models. Trends in Ecology & Evolution 33: 790-802. URL
490   https://doi.org/10.1016/j.tree.2018.08.001 .
491

Functions found in *AlleleShift* and their short description.

1   **Table 1.** Functions found in *AlleleShift* and their short descriptions

Function	Description
Preparation	
`VIF.subset`	Reduce the number of explanatory variables through Variance Inflation Factor analysis, with an option to plot a correlation matrix (`GGally::ggcorr`). Internally, the function calls the `BiodiversityR::ensemble.VIF.dataframe` function.
`environmental.novel`	Identify populations with some values of the explanatory variables that are outside the range of values used for calibration.
Analysis	
`count.model`	Calibration of RDA model (`vegan::rda`) with baseline allele counts as response and baseline bioclimatic variables as explanatory variables
`count.pred`	Prediction of allele counts (`vegan::predict.rda`) from explanatory variables. Explanatory variables correspond to the baseline climate to check the calibration
`freq.model`	Calibration of GAM model (`mgcv::gam`) with baseline allele frequencies as response and predicted baseline counts from the RDA model as explanatory variables
`freq.pred`	Prediction of allele frequencies (`mgcv::predict.gam`) from the predicted alleles counts of `count.pred`
`amova.rda`	Perform AMOVA with the outputs from RDA. The function returns an output that is similar to the output of `poppr::poppr.amova` so that results can be readily compared
Visualization	
`population.shift`	Shifts of populations in environmental space, with superellipses (`ggforce::geom_mark_ellipse`) and arrows between baseline and changed positions to show climatic shifts. Internally, the function calls vegan::ordiplot and various helper functions from BiodiversityR (Kindt 2020a) are used
`freq.ggplot`	Plots of baseline allele frequencies against predicted allele frequencies. Data points can be coloured differently by population or by allele
`shift.dot.ggplot`	Shifts of Allele Frequencies as Response to Climate Change
`shift.pie.ggplot`	Shifts of Allele Frequencies as Response to Climate Change via `ggforce::geom_arc_bar`
`pie.baker`	Helper function to prepare data for `shift.pie.ggplot` from the output of `freq.pred`.
`shift.moon.ggplot`	Shifts of Allele Frequencies as Response to Climate Change via `gggibbous::geom_moon`
`moon.waxer`	Helper function to prepare data for `shift.moon.ggplot` from the output of `freq.pred`

Function	Description
`shift.waffle.ggplot`	Shifts of Allele Frequencies as Response to Climate Change. Graphics are similar to `waffle::waffle`, but the graph is made *de novo* in *AlleleShift*
`waffle.baker`	Helper function to prepare data for `shift.waffle.ggplot` from the output of `freq.pred`
`shift.surf.ggplot`	Shifts of Allele Frequencies as Response to Climate Change, plotted in geographical space through smoothed regression surfaces (`vegan::ordisurf`)

2   **Notes**

3   *BiodiversityR* (Kindt & Coe, 2005; version 2.12-3 used in this manuscript); *ggforce* (Pedersen,
4   2000; version 0.3.2); *gggibbous* (Bramson 2019; version 0.1.0); mgcv (Wood, 2004; version 1.8-
5   31); *poppr* (Kamvar, Tabima & Grünwald, 2014; version 2.8.6); vegan (Oksanen et al., 2020;
6   version 2.5-6).

Output from function freq.pred for the changed climate for four populations and 2 alleles (A and B).

This is a subset of the result with all populations and all alleles for the example data sets in *AlleleShift* for *Populus trichocarpa.*

1 **Table 2.** Output from function `freq.pred` for the changed climate for four populations and 2
2 alleles (A and B). This is a subset of the result with all populations and all alleles for the example
3 data sets in *AlleleShift* for *Populus trichocarpa*.

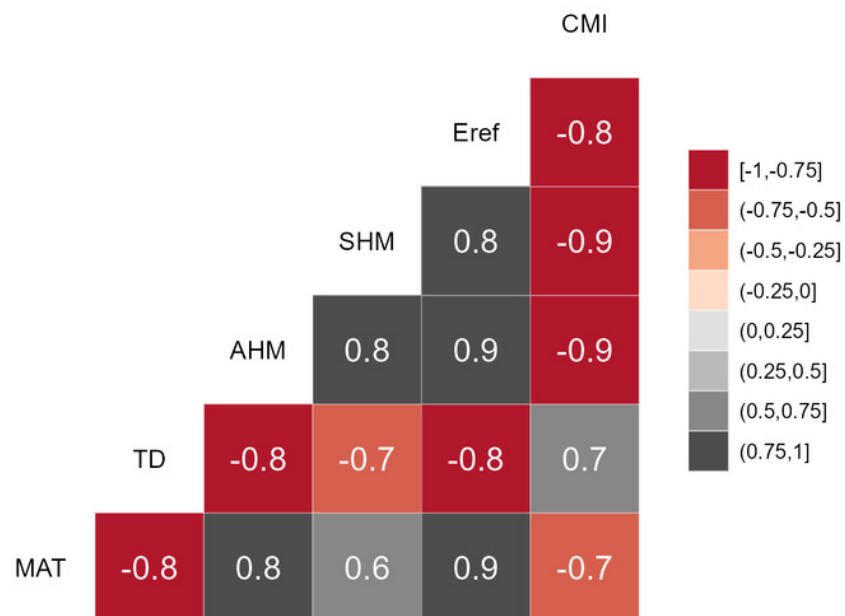
Population	Puyallup	Tahoe	Skagit	Nisqually
N	372	24	326	28
Locus	X01_10838495			
Allele.freq	0.172	0.625	0.236	0.107
A	64	15	77	3
B	308	9	249	25
Ap	144.376	23.471	75.381	15.165
Bp	227.624	0.529	250.619	12.835
N.e1	372	24	326	28
Freq.e1	0.388	0.978	0.231	0.542
Freq.e2	0.425	0.959	0.192	0.633
LCL	0.306	0.794	0.175	0.452
UCL	0.544	1.000	0.210	0.813
increasing	TRUE	TRUE	FALSE	TRUE
Locus	X01_16628872			
Allele.freq	0.169	0.083	0.181	0.250
A	63	2	59	7
B	309	22	267	21
Ap	38.751	0.509	52.668	-0.038
Bp	333.249	23.491	273.332	28.038
N.e1	372	24	326	28
Freq.e1	0.104	0.021	0.162	-0.001
Freq.e2	0.104	0.037	0.183	0.027
LCL	0.091	0.029	0.161	0.018
UCL	0.117	0.045	0.205	0.036
increasing	FALSE	FALSE	TRUE	FALSE

5 **Notes**
6 N, number of alleles sampled from population; Allele.freq, frequency of the minor allele in the
7 baseline climate calculated via `adegenet::make.freq`; A/B, count of the alleles in the
8 baseline climate; Ap/Bp, predicted counts of the alleles in the changed climate; N.e1, predicted
9 number of alleles; Freq.e1: predicted frequency of the minor allele via RDA; Freq.e2, predicted
10 frequency of the minor allele via GAM; LCL/UCL: lower/ upper confidence limits; increasing,
11 flag whether frequency is increasing.

Figure 1

Correlation matrices.

(A) Correlation matrix plot for the final subset of bioclimatic variables selected by the VIF.subset function. CMI: climatic moisture index, Eref: reference atmospheric evaporative demand, SHM: summer heat-to-moisture index, AHM: annual heat-to-moisture index, TD: continentality, MAT: mean annual temperature. (B) Correlation matrix plot for the minor alleles.

A

B

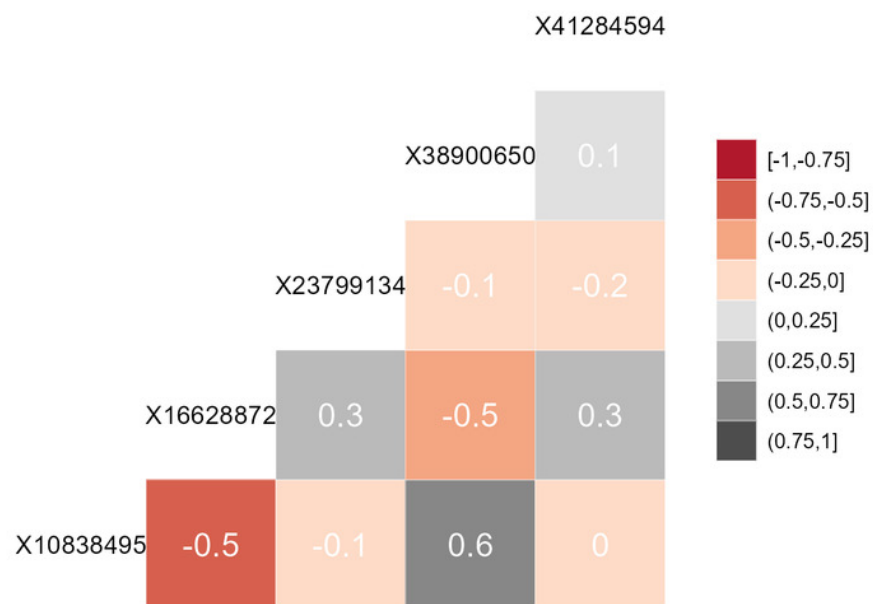



Figure 2

Ordination graphs.

(A) Principal Component Analysis (PCA) ordination graph for shifts in populations in environmental space. (B) Redundancy Analysis (RDA) ordination graph for shifts in populations with climate (a categorical variable with 'baseline' and 'future' as levels) as explanatory variable.

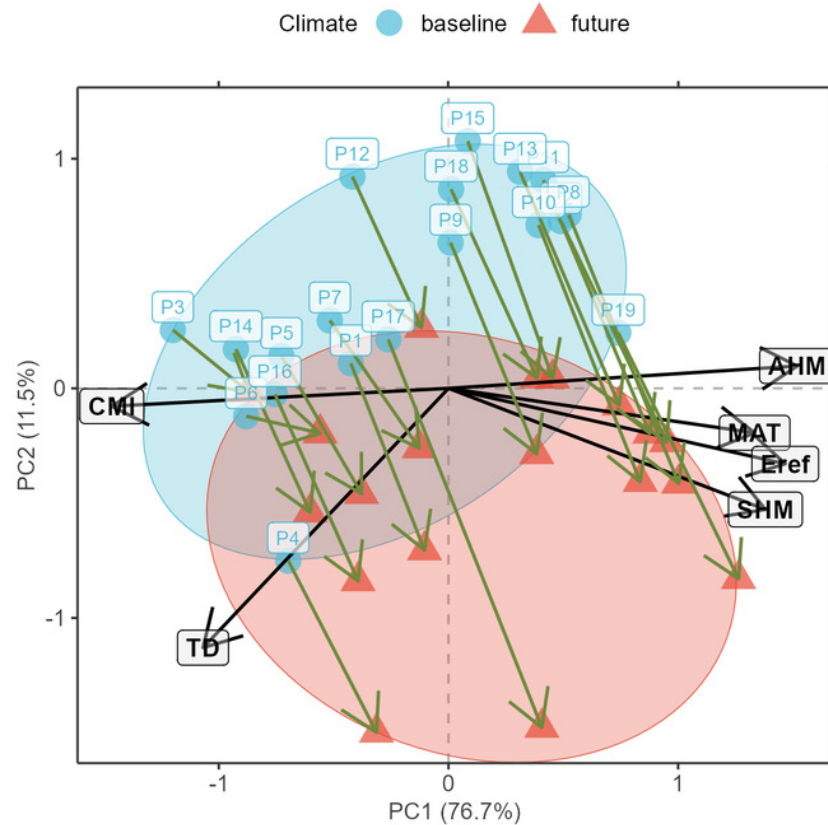
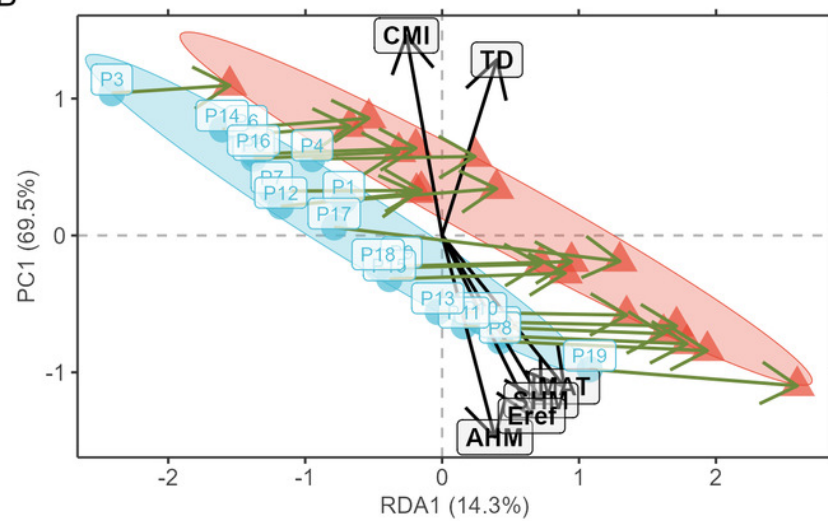



Figure 3

Plot of the actual frequency of the minor allele against the predicted frequency for the calibration data.

The 'olivegreen' reference lines indicate 1:1 (bold line), 1:1.05 and 1:0.95 (dashed lines) relationships. (A) Data for populations where a linear model explains more than 50% in allele frequencies. (B) Data for populations where a linear model explains less than 50% in allele frequencies.

A

Population : R^2

● Chilliwack (P1) : 0.95
● Columbia (P2) : 0.82
● Harrison (P4) : 0.90
● Nisqually (P8) : 0.54
● Nooksack (P9) : 0.84
● Olympic_Penninsula (P10) : 0.98
● Puyallup (P11) : 0.99
● Skagit (P13) : 0.55
● Skwawka (P14) : 0.90
● Skykomish (P15) : 0.89
● Squamish (P16) : 0.98
● Tahoe (P17) : 0.99
● Vancouver I. East (P18) : 0.59
● Willamette (P19) : 0.91

B

Population : R^2

● Dean (P3) : 0.15
● Homathko (P5) : 0.17
● Kitimat (P6) : 0.41
● Klinaklini (P7) : 0.42
● Salmon (P12) : 0.34

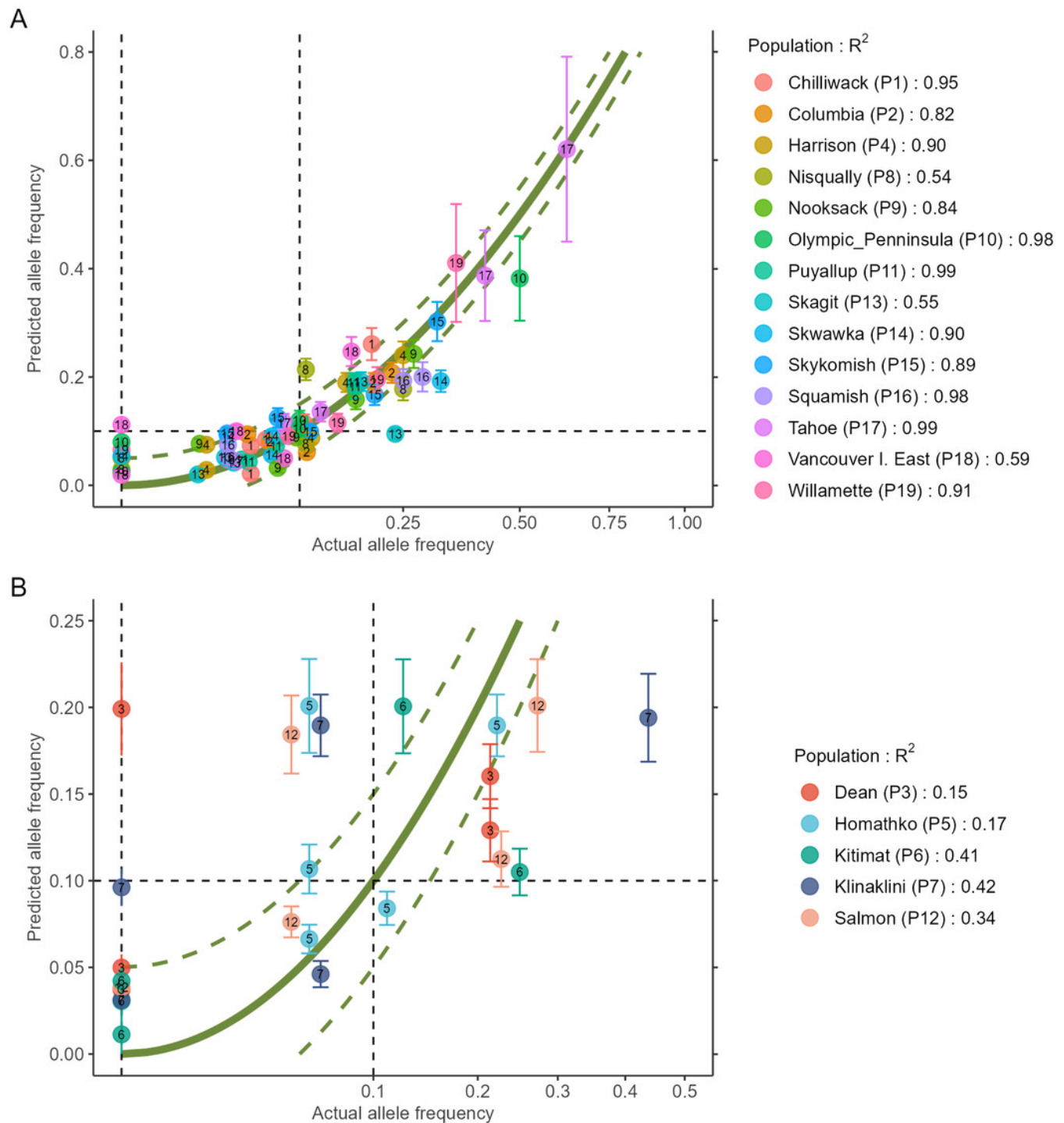



Figure 4

Depiction of changes in minor allele frequencies via AlleleShift::shift.dot.ggplot.

Black circles reflect baseline frequencies, blue circles future frequencies and vertical lines the confidence interval

Future change in allele frequencies ▬ decreasing ▬ increasing

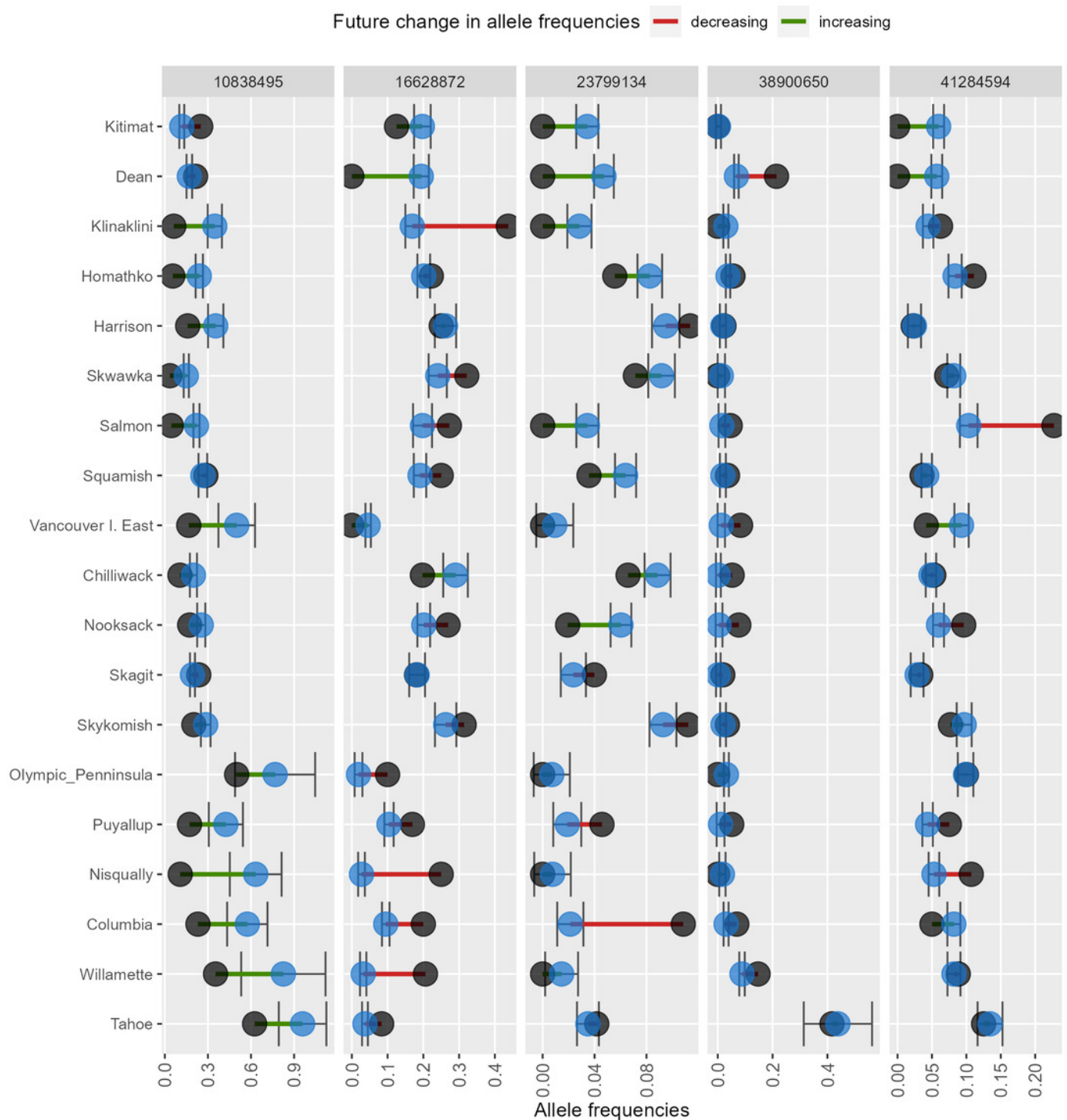



Figure 5

Depiction of changes in allele frequencies via AlleleShift::shift.pie.ggplot.

Columns on the left reflect baseline frequencies with frequency of the minor allele in black. Columns on the right reflect future frequencies, with colour of the arc and colour of the central circle reflecting frequencies and trends (red = decreasing, green = increasing) of the minor allele.

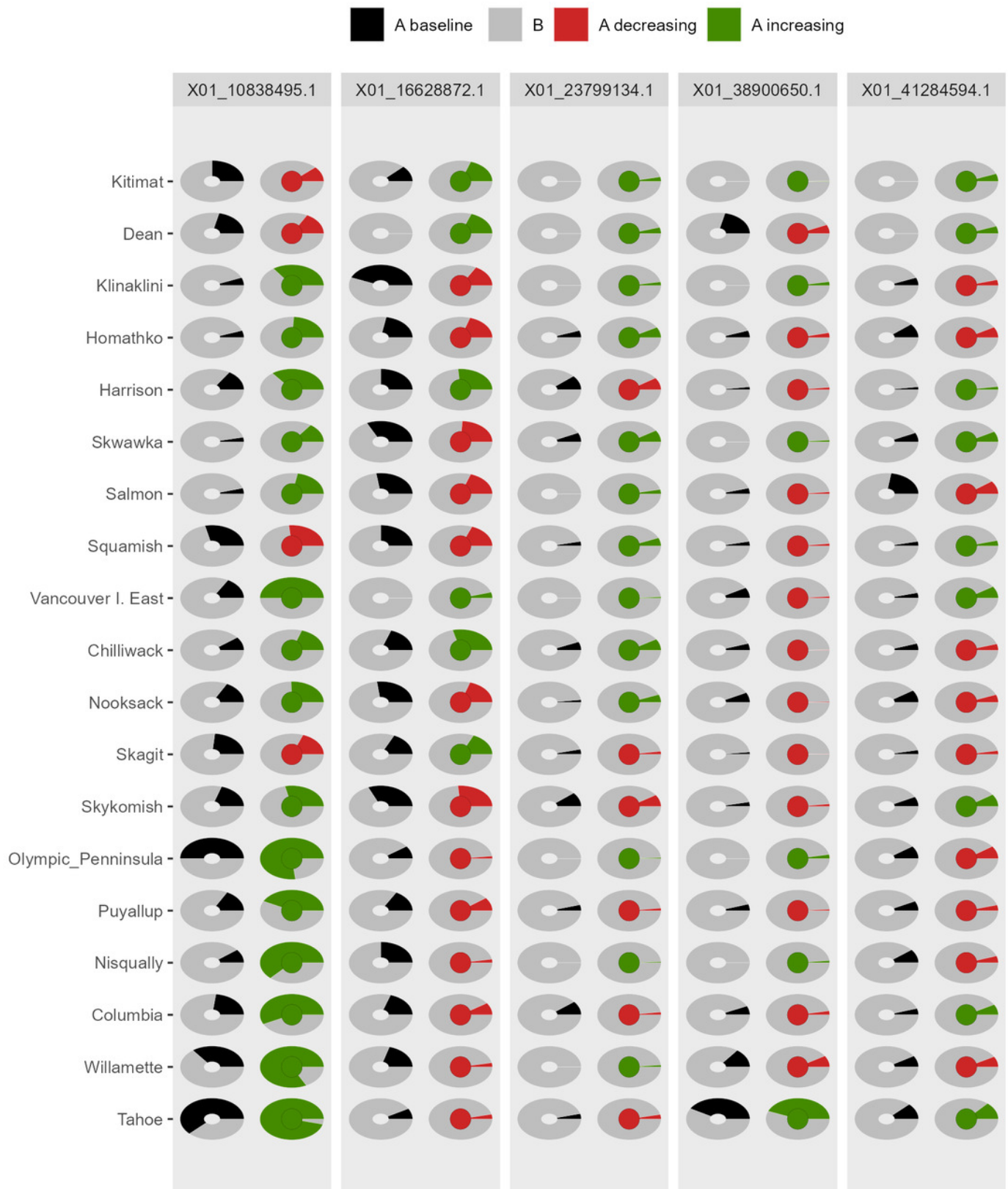



Figure 6

Depiction of changes in allele frequencies via AlleleShift::shift.moon.ggplot.

Columns on the left reflect baseline frequencies with frequency of the minor allele in white. Columns on the right reflect future frequencies, with colour of the waxing moon and colour of the central circle reflecting frequencies and trends (red = decreasing, green = increasing) of the minor allele.

Figure 7

Depiction of changes in allele frequencies via AlleleShift::shift.waffle.ggplot.

Each 'waffle' has 100 'cells'. Columns on the left reflect baseline frequencies with frequency of the minor allele in black. Columns on the right reflect future frequencies of the minor allele, with colour indicating trends (red = decreasing, green = increasing)

Figure 8

Depiction of changes in allele frequencies in geographical space via
AlleleShift::shift.surf.ggplot.

(A) Frequencies in the baseline climate for the minor allele for locus 10838495. (B)

Frequencies in a future climate for the same allele as in (A). (C) Frequencies in the baseline

climate for the minor allele for locus 23799134. (D) Frequencies in a future climate for the

same allele as in (D). Sizes of circles reflect the frequencies for the populations. Colours for

the future frequencies (B, D)_indicate trends (red = decreasing, green = increasing).

A

X01_10838495.1

(0.10, 0.15]
(0.15, 0.20]
(0.20, 0.25]
(0.25, 0.30]
(0.30, 0.35]
(0.35, 0.40]
(0.40, 0.45]
(0.45, 0.50]
(0.50, 0.55]

B

X01_10838495.1

(0.05, 0.10]
(0.10, 0.15]
(0.15, 0.20]
(0.20, 0.25]
(0.25, 0.30]
(0.30, 0.35]
(0.35, 0.40]
(0.40, 0.45]
(0.45, 0.50]
(0.50, 0.55]

C

X01_23799134.1

(0.000, 0.005]
(0.005, 0.010]
(0.010, 0.015]
(0.015, 0.020]
(0.020, 0.025]
(0.025, 0.030]
(0.030, 0.035]
(0.035, 0.040]
(0.040, 0.045]
(0.045, 0.050]
(0.050, 0.055]
(0.055, 0.060]
(0.060, 0.065]

D

X01_23799134.1

(0.030, 0.035]
(0.035, 0.040]
(0.040, 0.045]
(0.045, 0.050]
(0.050, 0.055]
(0.055, 0.060]
(0.060, 0.065]
(0.065, 0.070]