

Signals of Selection in UK Bumblebees

Natasha Ramsden¹

Supervised by Dr Peter Graystock¹

¹Imperial College London

1 Keywords

2 Genomics, bioinformatics, bumblebees, big data, HPC, selection.

3 2 Introduction

4 Bumblebees are critical in crop and wildflower pollination globally: these charismatic species are there-
5 fore vitally important economically and ecologically; for food security and ecosystem stability (Goulson
6 et al., 2008; Cameron and Sadd, 2020; Sun et al., 2021). However, many species of bumblebee (*Bom-*
7 *bus sp.*) face population declines caused by multifactorial stressors. Habitat loss, fragmentation and
8 degradation; climate change; pathogens and pesticides have worked in combination to drive the decline
9 of some species (Goulson et al., 2008; Cameron and Sadd, 2020). Understanding underlying genetic
10 factors linked to these trends will be important in making inferences about population health and future
11 trajectories, as well as implementing successful conservation action.

12 This study will focus on three bumblebee species which are resident in the United Kingdom: *Bombus*
13 *terrestris*, *Bombus hortorum* and *Bombus ruderatus*. Whilst *B. terrestris* and *B. hortorum* are widespread
14 across the UK, *B. ruderatus* has faced historic population decline and has a limited and fragmented
15 geographical range (Ellis et al., 2005). Differences in the population dynamics of these three species
16 could be explained by scrutiny of their genomes to understand differences in the selective events they
17 face.

18 Using restriction-site associated DNA sequencing (RADSeq) data previously collected from these
19 species at multiple sites in the UK, I aim to detect loci under positive selection. Selection in these
20 species of bumblebee has not previously been investigated, so this novel study may be essential in
21 understanding the diverging population trends and in targeting conservation plans. Specifically I aim
22 to detect loci under selection and to identify differences in these findings between the species. Where
23 possible, I will functionally annotate the selected sites by identifying genes associated with the significant
24 loci. Finally I will attempt to identify the functional categories that these genes belong to in order to
25 understand the underlying drivers and consequences of these selective forces.

26 3 Proposed Methods

27 In order to achieve the main goals outlined above I will undertake the following methods:

28 1. I will implement the *STACKS* pipeline (Catchen et al., 2013) for genome assembly and sample
29 filtering of the raw RADSeq data (Rochette and Catchen, 2017).

30 2. Following this, I intend to execute multiple methods for identifying loci under positive selection in
31 the *Bombus sp.* including the commonly-used software *BAYESCAN* (Foll and Gaggiotti, 2008; Ahrens
32 et al., 2018) (e.g. Blanco-Bercial and Bucklin, 2016; Kang et al., 2017; Leiva et al., 2019; de Jong et al.,
33 2021). Loci determined to be under positive selection will be compared to those identified using different
34 methods, in order to isolate those that are consistently identified as outliers. This will improve robustness
35 of inferences by reducing the impact of false positives. Through this analysis I will test the hypothesis
36 that RADSeq data can be used to detect selection in non-model species.

37 3. A comparison of these significant loci between the species will be made to determine whether
 38 the species' genomes are similarly or differently affected by selection. I hypothesise that the species in
 39 decline (*B. ruderatus*) will display different signatures of selection to the other species.
 40 4. Since reference genomes are available for *B. terrestris* and *B. hortorum*, I will place the significant
 41 loci within the physical context of their genomes to identify genes in linkage with these loci, which may
 42 therefore be affected by selection (Manel et al., 2016). I would expect to find signals of selection in areas
 43 of the genome which are functionally related to the stressors that these species face.

44 **4 Anticipated Outputs and Outcomes**

45 From the comparison of selection-detection methods I would hope to identify several loci consistently
 46 established to be under selection in the *Bombus sp.*' genomes. By comparing which loci were detected
 47 I hope to identify any common patterns between the species, which could be indicative of selection
 48 pressures shared by the genus; as well as any differences, which may offer insight into the differing
 49 population trends of the individual species. Where possible, I would like to identify any relevant functional
 50 roles associated with genes in linkage with the detected loci to put this study into some environmental
 51 context.

52 **5 Timeline**

Table 1: Gantt chart of expected timeline. Numbered tasks indicate the broad methodological sections outlined above.

	April	May	June	July	Aug	Sept
Method						
Task 1: Filtering and assembly						
Task 2: Detecting selection						
Task 3: Species comparison						
Task 4: Functional categorisation						
Write-up						
Introduction						
Methods (simultaneous to tasks)						
Results						
Discussion						
Viva prep						

53 **6 Budget**

54 8Tb External hard-drive to backup data and run some analyses locally (£200) and 21" monitor to improve
 55 data exploration (£230).

References

- Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L. and Dudaniec, R. Y. (2018), 'The search for loci under selection: trends, biases and progress', *Molecular Ecology* **27**(6), 1342–1356.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.14549>
- Blanco-Bercial, L. and Bucklin, A. (2016), 'New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*', *Molecular Ecology* **25**(7), 1566–1580.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13581>
- Cameron, S. A. and Sadd, B. M. (2020), 'Global Trends in Bumble Bee Health', *Annual Review of Entomology* **65**(1), 209–232.
URL: <https://doi.org/10.1146/annurev-ento-011118-111847>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. and Cresko, W. A. (2013), 'Stacks: an analysis tool set for population genomics', *Molecular Ecology* **22**(11), 3124–3140.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.12354>
- de Jong, M. J., Lovatt, F. and Hoelzel, A. R. (2021), 'Detecting genetic signals of selection in heavily bottlenecked reindeer populations by comparing parallel founder events', *Molecular Ecology* **30**(7), 1642–1658.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15837>
- Ellis, J. S., Knight, M. E. and Goulson, D. (2005), 'Delineating species for conservation using mitochondrial sequence data: the taxonomic status of two problematic *Bombus* species (Hymenoptera: Apidae)', *Journal of Insect Conservation* **9**(2), 75–83.
URL: <http://link.springer.com/10.1007/s10841-004-4168-0>
- Foll, M. and Gaggiotti, O. (2008), 'A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective', *Genetics* **180**(2), 977–993.
URL: <https://doi.org/10.1534/genetics.108.092221>
- Goulson, D., Lye, G. and Darvill, B. (2008), 'Decline and Conservation of Bumble Bees', *Annual Review of Entomology* **53**(1), 191–208.
URL: <https://doi.org/10.1146/annurev.ento.53.103106.093454>
- Kang, J., Ma, X. and He, S. (2017), 'Population genetics analysis of the Nujiang catfish *Creteuchiloglanis macropterus* through a genome-wide single nucleotide polymorphisms resource generated by RAD-seq', *Scientific Reports* **7**(1), 2813.
URL: <https://www.nature.com/articles/s41598-017-02853-3>
- Leiva, C., Taboada, S., Kenny, N. J., Combosch, D., Giribet, G., Jombart, T. and Riesgo, A. (2019), 'Population substructure and signals of divergent adaptive selection despite admixture in the sponge *Dendrilla antarctica* from shallow waters surrounding the Antarctic Peninsula', *Molecular Ecology*

92 **28**(13), 3151–3170.

93 **URL:** <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15135>

94 Manel, S., Perrier, C., Pratlong, M., Abi-Rached, L., Paganini, J., Pontarotti, P. and Aurelle, D. (2016),

95 ‘Genomic resources and their influence on the detection of the signal of positive selection in genome

96 scans’, *Molecular Ecology* **25**(1), 170–184.

97 **URL:** <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13468>

98 Rochette, N. C. and Catchen, J. M. (2017), ‘Deriving genotypes from RAD-seq short-read data using

99 Stacks’, *Nature Protocols* **12**(12), 2640–2659.

100 **URL:** <https://www.nature.com/articles/nprot.2017.123>

101 Sun, C., Huang, J., Wang, Y., Zhao, X., Su, L., Thomas, G. W. C., Zhao, M., Zhang, X., Jungreis, I.,

102 Kellis, M., Vicario, S., Sharakhov, I. V., Bondarenko, S. M., Hasselmann, M., Kim, C. N., Paten, B.,

103 Penso-Dolfin, L., Wang, L., Chang, Y., Gao, Q., Ma, L., Ma, L., Zhang, Z., Zhang, H., Zhang, H.,

104 Ruzzante, L., Robertson, H. M., Zhu, Y., Liu, Y., Yang, H., Ding, L., Wang, Q., Ma, D., Xu, W., Liang,

105 C., Itgen, M. W., Mee, L., Cao, G., Zhang, Z., Sadd, B. M., Hahn, M. W., Schaack, S., Barribeau,

106 S. M., Williams, P. H., Waterhouse, R. M. and Mueller, R. L. (2021), ‘Genus-Wide Characterization of

107 Bumblebee Genomes Provides Insights into Their Evolution and Variation in Ecological and Behavioral

108 Traits’, *Molecular Biology and Evolution* **38**(2), 486–501.

109 **URL:** <https://doi.org/10.1093/molbev/msaa240>

I have seen and approved the proposal and the budget.

A handwritten signature in black ink, appearing to read 'P. Graystock', with a stylized flourish at the end.

10/04/2022

Dr Peter Graystock