Problem:

ProAska for translocation has only been sourced from Strickland West

Question:

Which additional sites should be sampled from to improve the diversity and representativeness of the translocation population?

How to answer:

Determine where unique diversity is seen in the species and sample from sites which capture this diversity.

Method

1. Subset data to individuals of interest (Analysis.1)
2. Filter to maximum locus missingness of 30% and one snp per locus (important)
3. Calculate diversity statistics.
4. For each site, create a list of alleles found at >=5% frequency
   1. Determine how many total alleles per site
   2. Determine how many private alleles per site (alleles not found elsewhere)

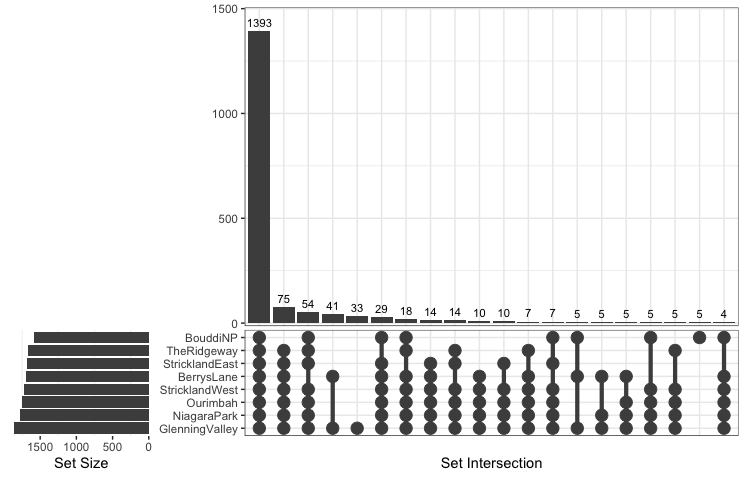
At this stage we found that all sites share the majority of common alleles (72%) and that Strickland West specifically contains 89% of the common alleles in ProAska. Most sites did not have many private alleles, however, Glenning Valley had notably more private alleles than the other sites. Together, the samples from Glenning Valley and Strickland West contain 98% of all common alleles observed in ProAska, therefore we recommend more samples are taken from Glenning Valley to improve the diversity of the translocation collection.

Summary

* Strickland west collections potentially already contain a large proportion of the species common allele diversity. To improve this further, we recommend sampling from the Glenning Valley site which contains unique diversity.

Other notes:

* Bouddi NP has lower heterozygosity and contains fewer of the common alleles than the other populations – may be genetically depleted because of isolation and drift?

Figure 1: plain upset plot summary of common alleles (>=5% frequency in ProAska) shared between sites.

The first column shows all sites share 1393 alleles. The second column shows that all sites excluding Bouddi NP share an additional 75 alleles.

Bouddi has the fewest alleles (set size) and Glenning Valley has the most alleles.

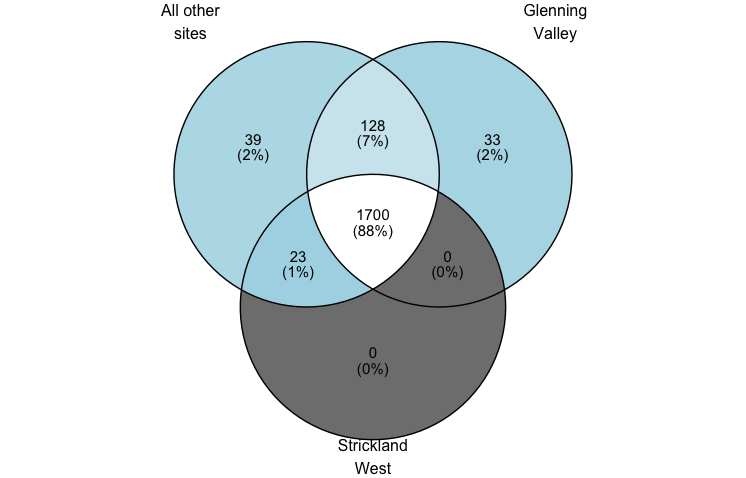


Figure 2: Venn diagram of common alleles shared between Glenning Valley, Strickland West, and all other ProAska sites. Together GV and SW contain 98% of the observed common alleles.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| site | ar | obs\_het | exp\_het | fis | loci | n | PA | TA |
| BerrysLane | 1.466 | 0.232 | 0.247 | 0.015 | 962 | 3 | 1 | 1690 |
| BouddiNP | 1.235 | 0.119 | 0.122 | -0.005 | 962 | 6 | 5 | 1590 |
| GlenningValley | 1.657 | 0.275 | 0.346 | 0.186 | 962 | 7 | 33 | 1861 |
| NiagaraPark | 1.461 | 0.191 | 0.234 | 0.164 | 962 | 12 | 1 | 1781 |
| Ourimbah | 1.472 | 0.21 | 0.233 | 0.09 | 962 | 8 | 0 | 1755 |
| StricklandEast | 1.424 | 0.198 | 0.209 | 0.027 | 962 | 4 | 0 | 1678 |
| StricklandWest | 1.432 | 0.192 | 0.214 | 0.084 | 962 | 7 | 0 | 1723 |
| TheRidgeway | 1.448 | 0.227 | 0.219 | -0.066 | 962 | 3 | 0 | 1670 |
| Total |  |  |  |  |  |  |  | 1923 |

PA = Private alleles

TA = Total alleles