word count abstract: 86, main text with legends, excluding references: 952

Development of a large set of diagnostic SNP markers using ddRAD-seq to study hybridization in *Rhinanthus major* and *R. minor* (Orobanchaceae)

Khaled Mirzaei & Renate A. Wesselingh*

Biodiversity Research Centre, Earth and Life Institute, UCLouvain, Croix du Sud 4 box L7.07.04, B-1348 Louvain-la-Neuve, Belgium

*Corresponding author, e-mail: renate.wesselingh@uclouvain.be, ORCID: 0000-0003-0241-2255

Acknowledgements

This is publication no. BRC362 of the Biodiversity Research Centre of UCLouvain. We thank Bertrand Bearzatto, Vinciane Mossion and Marie Eve Renard for their help during the development of the ddRADseq protocol for *Rhinanthus*, Vanessa Brys and Wim Meert at the Genomics Core KULeuven for technical assistance with the Blue Pippin and the Fund for Scientific Research-FNRS (Research Credit J000614F) and the Fonds Spéciaux de Recherche of UCLouvain (FSR project 2016) for financial support. Computational resources have been provided by the supercomputing facilities of UCLouvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region.

Declarations

Funding see Acknowledgements.

Data availability All raw and processed data are kept at UCLouvain and can be shared upon request, information on the diagnostic loci and their SNPs is provided as online supplementary material.

Animal Research (Ethics) NA

Consent to Participate (Ethics) NA

Consent to Publish (Ethics) NA

Plant Reproducibility Samples were taken in the wild (not from pure lineages) and the species in question are annuals, so the plants themselves cannot be reproduced. But the leaf samples are kept at –80°C at UCLouvain and could be reused for further analysis.

Clinical Trials Registration NA

Authors contribution KM did all the practical work and data analyses and wrote the first version of the manuscript, RAW assisted in discussions and revisions of the manuscript.

Conflict of Interest None

Abstract

Rhinanthus major and *R. minor* are two annual plant species that are known to hybridize in nature and in which unilateral introgression is likely to occur. Here we used double-digest restriction site-associated DNA sequencing technology (ddRAD-seq) to detect 16,932 genome-wide SNPs in *R. major* and *R. minor* individuals from 9 populations. After screening these SNP markers across both species, we selected 1106 putative loci that contain diagnostic, species-specific SNPs, which can be used for assessing and monitoring hybridization and introgression between *R. major* and *R. minor*.

Key words: ddRAD-seq, diagnostic markers, SNP, *Rhinanthus*

Hybridization and introgression are evolutionary phenomena that take place in around 25% of plant species and are likely to be one of the most important sources of evolution and domestication of flowering plants (Baack and Rieseberg 2007). When hybridization and especially introgression occur, accurate identification of species can be significantly impaired (Mallet 2005; Soltis and Soltis 2009).

Rhinanthus major Ehrh. (syn. *R. angustifolius* C.C. Gmelin) and *R. minor* L. (both 2n = 2x = 22: Hambler (1954)) are root hemiparasitic annual plants occurring in a diverse range of grassland habitats throughout northern and central Europe. Wherever agricultural practices have been intensified, they have become less common and are often only found in nature reserves. Because they parasitize dominant grass species, they are important tools in the restauration of species-rich grasslands (Westbury et al. 2006, Těšitel et al. 2018). Their ecological niches overlap, so the species can co-occur (Westbury 2004; Ducarme et al. 2010). Only a few morphological features that can be used for distinguishing the two species, including bract color and corolla shape and size (Kwak 1978). Although some prezygotic barriers against hybridization exist (Natalis and Wesselingh 2012), hybrid formation occurs readily and F_1 hybrids perform as well as the parental species (Wesselingh et al. 2019). In later generations, an overall pollinator preference for *R. major* (Natalis and Wesselingh 2013; Ducarme and Wesselingh 2013) causes asymmetric introgression from *R. minor* into *R. major* (Ducarme et al. 2010). To study introgression in detail, we needed to greatly expand the limited set of markers we used previously (Ducarme & Wesselingh 2005, Ducarme et al. 2008).

We used a total of 76 samples previously collected in the field or grown from seed in the greenhouse from our two seed source populations for hybrid production (DB and HO). The 25 *Rhinanthus major* samples came from nature reserve Doode Bemde (province of Flemish Brabant). For *Rhinanthus minor* we used 51 samples, of which 26 came from two pure

populations: 11 from nature reserve Housta (HO; Braine-le-Château, province of Walloon Brabant), and 15 from Jardin Massart (JM; Auderghem, Brussels, population founded with seeds from Chantemelle, province of Luxemburg). We also included samples from mixed populations that contained none of the *R. major* markers and thus were pure *R. minor*: 4 from Karthuizerduinen (KD; Nieuwpoort, province of West Flanders), 5 from Ter Yde (TY, Oostduinkerke, province of West Flanders), 4 from Kalkense Meersen (KM; Uitbergen, province of East Flanders), 3 from Kijkverdriet (KV; Ravels, province of Antwerpen), 5 from Achter Schoonhoven (AS; Schoonhoven, province of Flemish Brabant) and 4 from Messelbroek (MB; province of Flemish Brabant).

DNA was extracted from young leaves using the CTAB protocol (Doyle and Doyle 1987) and the ddRAD library was prepared based on the protocol described by Peterson et al. (2012). We tested different combinations of restriction enzymes and chose *Eco*RI and *Nla*III for double digestion. The ligated fragments were size-selected (325–425bp) using a Blue Pippin (BDF2010, Cassette type: 2% DF Marker Q2). Illumina indexes were added through PCR and the pooled libraries were paired-end sequenced (PE150) on the Illumina Hi-Seq 2500 platform at the Genomics Core of KULeuven and with Novaseq at Genewiz. Quality control checks on raw sequence data were performed using FastQC (Andrews 2010). Filtering, demultiplexing and SNP detection were done with Stacks v. 2.1 (minimum stack depth = 5, distance between stacks = 3, distance between catalog loci = 3; Catchen et al. 2013).

In total, 402,520,746 raw sequencing reads with high quality were generated from the 76 samples with an average of 5,296,325 reads per individual. Depth of coverage for the processed samples varied from 8x to 20x, with an average of 12x per individual. We initially obtained 11,171 loci that were shared by at least 70% of all individuals. After removing loci that only contained SNPs with a minor allele frequency below 0.05, 8,545 loci with 16,932 SNPs in total were retained for further analysis. Of these, 57.7% were confirmed to be

transitions, and the observed transition-transversion ratio between the two species was 1.36. Principal component analysis (PCA, performed with Tassel 5; Bradbury et al. 2007) of all discovered SNP markers divided the nine populations into four groups, with a clear and strong differentiation between *R. major* and *R. minor* (Figure 1).

Genetic diversity analysis performed in STACKS (Catchen et al. 2013) showed that the number of private alleles and the observed heterozygosity were higher in *R. major* than in *R. minor* (Table 1). The fixation index (F_{ST}) between the two species was 0.51, and a higher inbreeding coefficient was found for *R. minor* (0.22) compared to *R. major* (0.0468). This is in line with the differences in breeding system between the two species. Ducarme and Wesselingh (2013) found an average outcrossing rate of 0.76 for *R. major*, while this was only 0.13 for *R. minor*.

We selected a subset of 1106 loci with 1–3 species-specific SNPs (1800 SNPs in total) that were fixed for a single allele in *R. minor* and carried one or two alternate alleles in *R. major* and as our diagnostic loci (Online Resource 1). These can now be used to study introgression in this species pair in great detail in mixed populations in Belgium and neighbouring countries, based on phylogeographic patterns found in Europe (Vrancken et al. 2012).

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17:513–518. https://doi.org/10.1016/j.gde.2007.09.001

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15.

Ducarme V, Risterucci AM, Wesselingh RA (2008) PERMANENT GENETIC RESOURCES: Development of microsatellite markers in *Rhinanthus angustifolius* and cross-species amplification. Mol Ecol Resour 8:384–386. https://doi.org/10.1111/j.1471-8286.2007.01966.x

Ducarme V, Vrancken J, Wesselingh RA (2010) Hybridization in annual plants: patterns and dynamics during a four-year study in mixed *Rhinanthus* populations. Folia Geobot 45:387–405. https://doi.org/10.1007/s12224-010-9077-2

Ducarme V, Wesselingh RA (2013) Outcrossing rates in two self-compatible, hybridising *Rhinanthus* species: implications for hybrid formation. Plant Biol 15:541–547. https://doi.org/10.1111/j.1438-8677.2012.00668.x

Hambler DJ (1954) Cytology of the Scrophulariaceae and Orobanchaceae. Nature 174:838. https://doi.org/10.1038/174838a0

142 Kwak MM (1978) Pollination, hybridization and ethological isolation of *Rhinanthus minor*
143 and *R. serotinus* (*Rhinanthoideae*: *Scrophulariaceae*) by bumblebees (*Bombus* Latr.).
144 Taxon 27:145–158. https://doi.org/10.2307/1220235
145 Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237.
146 https://doi.org/10.1016/j.tree.2005.02.010
147 Natalis LC, Wesselingh RA (2012) Post-pollination barriers and their role in asymmetric
148 hybridization in *Rhinanthus* (Orobanchaceae). Am J Bot 99:1847–1856.
149 https://doi.org/10.3732/ajb.1200085
150 Natalis LC, Wesselingh RA (2013) Parental frequencies and spatial configuration shape
151 bumblebee behavior and floral isolation in hybridizing *Rhinanthus*. Evolution
152 67:1692–1705. https://doi.org/10.1111/evo.12044
153 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an
154 inexpensive method for de novo SNP discovery and genotyping in model and non-
155 model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135
156 Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol
157 60:561–588. https://doi.org/10.1146/annurev.arplant.043008.092039
158 Těšitel J, Mládek J, Fajmon K, Blažek P, Mudrák O, Hölzel N (2018) Reversing expansion of
159 *Calamagrostis epigejos* in a grassland biodiversity hotspot: Hemiparasitic *Rhinanthus*
160 *major* does a better job than increased mowing intensity. Appl Veg Sci 21:104–112.
161 https://doi.org/10.1111/avsc.12339
162 Vrancken J, Brochmann C, Wesselingh RA (2012) A European phylogeography of
163 *Rhinanthus minor* compared to *R. angustifolius*: unexpected splits and signs of
164 hybridization. Ecology and Evolution 2: 1531–1548. https://doi.org/10.1002/ece3.276

Wesselingh RA, Hořčicová Š, Mirzaei K (2019) Fitness of reciprocal F_1 hybrids between *Rhinanthus minor* and *Rhinanthus major* under controlled conditions and in the field. J Evol Biol 32:931–942. https://doi.org/10.1111/jeb.13492

Westbury DB (2004) Biological Flora of the British Isles. *Rhinanthus minor* L. J Ecol 92:906–927. https://doi.org/10.1111/j.0022-0477.2004.00929.x

Westbury DB, Davies A, Woodcock BA, Dunnett NP (2006). Seeds of change: the value of using *Rhinanthus minor* in grassland restoration. J Veg Sci 17:435–446. https://doi.org/10.1111/j.1654-1103.2006.tb02464.x

174 **Table 1.** Values of genetic diversity within species and populations. A_P = number of private
175 alleles, H_{obs} = observed heterozygosity, H_{exp} = expected heterozygosity, F_{IS} = inbreeding
176 coefficient.

Species	*N*	A_P	H_{obs}	H_{exp}	F_{IS}
R. major	25	6038	0.165±0.001	0.171±0.001	0.046±0.03
R. minor	51	4393	0.062±0.001	0.144±0.001	0.225±0.01

177

178

179

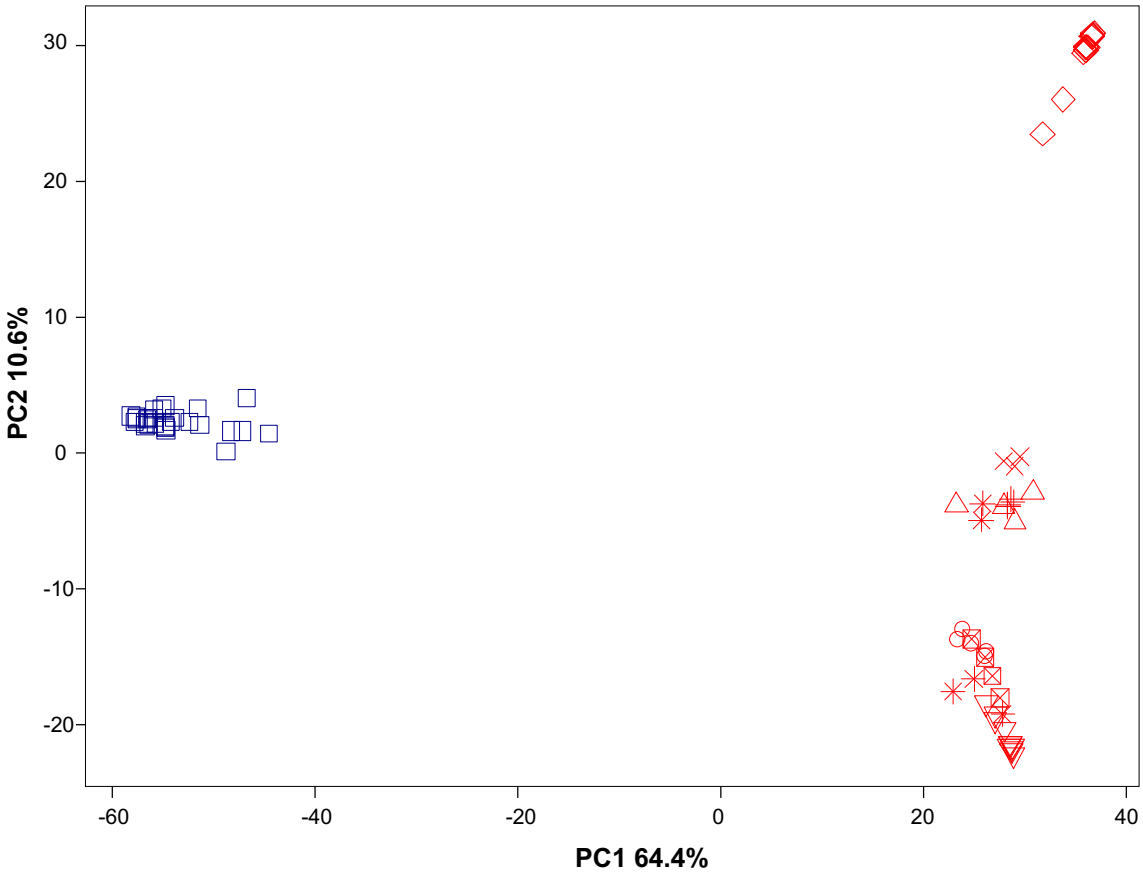

Figure 1. Principal component analysis (PCA) based on 16,932 SNP markers. *Rhinanthus minor* populations are indicated with red symbols: HO (▽), AS (○), MB (⊠), TY (✳), KD (Δ), KM (+), KV (×), JM (◇), and *R. major* samples (DB) by blue squares.

186 Electronic Supplementary Material 1

187

188 A list of 1106 loci with species-specific SNPs for *Rhinanthus minor* and *R. major*

189 Explanation of column headers in file ESM_1.csv

190 Locus = sequential locus number

191 Position = position of single nucleotide polymorphism in sequence

192 Allele_R.minor = nucleotide in *R. minor* sequence

193 Allele1_R.major = nucleotide in *R. major* sequence

194 Sequence_Rhinanthus_minor = the full sequence of the locus in *R. minor*