SUPPLEMENTARY DATA

MARKER INFORMATION

Identification of microsatellite markers and their ability to discriminate between *Cardamine amara* and *C. rivularis* auct.

Approach

454 sequencing

Total genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions from one individual of each of the assumed diploid progenitors, *C. amara* and *C. rivularis* auct., originating at Urnerboden. Sequencing libraries, sequencing of a ½ picotiter plate for each sample using the GS FLX instrument (Roche, Basel, Switzerland) and adapter trimming of the resulting data were performed at the Functional Genomics Center (Zürich, Switzerland).

Microsatellite loci identification

Microsatellite repeats were identified directly based on unassembled reads using msatcommander v 0.8.2 (Faircloth, 2008). SSR loci with a higher number of repeat units (dinucleotides containing at least 10, trinucleotides with at least 7, tetra- and pentanucleotides with minimum 6 repeat units) were searched for. Primers were designed directly during the repeat identification as implemented in msatcommander, under default settings.

Identification of species specificity in-silico

Our strategy was to target progenitor-specific loci. To identify the ability of the microsatellite markers to discriminate between the two *Cardamine* species, the primers designed for one species were compared with the sequences (reads) of the other species. At first sequence databases were created using the formatdb script as implemented in BLAST (Altschul *et al.*, 1990) from the *C. amara* and *C. rivularis* auct. reads. Afterwards the primer sequences of one species (combined to two fasta files, one containing all forward and the other all reverse primer sequences) were compared to the sequence database of the other species running the blastn function in BLAST. This comparison was performed for the forward and reverse primers separately. The search results were visualized using the PLAN web application (He *et al.*, 2007). Those markers for which at least one primer did not match any sequence of the other species at the last 3 bp at its 3' end were selected, reduced for duplicities (i.e., collapsing reads/markers with identical sequences into a single marker) and tested experimentally to confirm their specificity.

Cross-amplification tests

The markers selected based on BLAST comparisons were amplified on a subset of four individuals of each *C. amara* and *C. rivularis* auct. from Urnerboden. PCRs were performed in 25 µl reactions containing 1x PCR buffer with KCl (Fermentas, St. Leon, Germany), 2.5 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of both primers, 0.5 U of Taq DNA polymerase (Fermentas) and approximately 20 ng of DNA. The cycling conditions were: 5 min at 95°C followed by 35 cycles of 95°C for 30s, 55-58°C for 30s and 72°C for 1.5 min and final extension at 72°C for 10 min. The amplification products were visualized on 2.5% agarose gels stained with ethidium bromide. Those markers that did not cross-amplify to any accession of the other species were amplified again under the conditions modified according to Schuelke (2000), to facilitate fluorescent labeling of the PCR products. The samples were analysed on ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA, USA) with GeneScan 500 LIZ internal size standard (Applied Biosystems), in order to estimate the fragment lengths and quality (presence of unspecific products or potential

locus duplications). GeneMarker v. 1.80 (Softgenetics, State College, PA, USA) was used to call the allele sizes. The specificity of the markers was finally verified on the entire set of sampled individuals, checking the products on agarose gels.

Results

454 sequencing, microsatellite identification and BLAST comparisons

The 454 sequencing yielded 593,279 and 576,096 reads for *C. amara* and *C. rivularis* auct., respectively, with an average read length of 374 bp. For *C. amara* 4,542 reads containing microsatellite repeats were identified. In *C. rivularis* auct. 4,954 repeats contained microsatellite repeats. Primers could be designed for 777 and 673 repeats for *C. amara* and *C. rivularis* auct., respectively. Based on the BLAST analyses only 31 primer pairs designed for *C. amara* did not match significantly the *C. rivularis* auct. sequences. In case of *C. rivularis* auct., 29 primer pairs were shown to be potentially species-specific.

Characterisation of the microsatellite loci and experimental evaluation of their species specificity

Amplification of the markers selected according to the BLAST search results in four accessions of each species showed that only 15 loci of *C. amara* and 9 loci of *C. rivularis* auct. were not amplifying in the other species (see Table S1). Products of unspecific amplification were not observed in any of the markers. The extended amplification tests, however, revealed that three *C. rivularis* auct. markers (Criv2, Criv20 and Criv48) and one *C. amara* marker (Cama16) failed repeatedly to amplify in a high number of accessions (indicating the occurrence of null alleles) and/or produced poorly scorable peaks, and were therefore eliminated. Four markers showing initial specificity to *C. amara* were found to cross-amplify in *C. rivularis* auct. and *C. pratensis* as well (these loci were kept in the analyses and denoted as C-all loci in Table S1).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Faircloth BC. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8:** 92–94.

He J, Dai X, Zhao X. 2007. PLAN: A web platform for automating high-throughput BLAST searches and for managing and mining results. *BMC Bioinformatics* **8**: 53.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18:** 233–234.

Table S1. Characteristics of the microsatellite markers used in the present study. Markers excluded from the study due to the occurrence of null alleles and/or ambiguous allele determination are marked by asterisks

Marker	Repeat motif	Primer sequence (5'→3')	Allele size range (bp)	GenBank accession no
Criv loci (*C. rivularis* auct./*C. pratensis*-specific loci)				
Criv15	$(GTT)_7$	F: GCATTGTCGTAAGCTGCC R: CACAGCAAACCATGTGCAAC	218-262	KF998167
Criv16	$(AAC)_8$	F: CTGATTTGACAAGCCAAGGG R: TTGCTATTCGCCAGCTTCG	398-414	KF998168
Criv17	$(CTT)_7$	F: CTTCTTCGCTGGGAATCG R: CACGGAGAAGTCCAGTTGC	218-262	KF998169
Criv18	$(CATCT)_8$	F: CCATACTTGCGCCTCGTC R: TGCAAAGGCGTTAGGACAAC	237-247	KF998170
Criv23	$(CTT)_6(CGT)_3(CTT)_8$	F: AATTCATTTCTCCTTGCCGAC R: AGGCAACGAATCCGCTAAC	329-387	KF998172
Criv34	$(CTT)_7$	F: CCTGAGGCAATCATCCTT R: GGTGGTACACGTGATCTGC	180-205	KF998173
Criv2*	$(ATC)_5N_{41}(ATC)_9$	F: AACTTCTCAGGGCCGGTTC R: CCCTATCTCTCCATGTGGGC	146-159	KF998166
Criv20*	$(AAG)_8$	F: CCTCCCAAAGAATGCCCG R: TTGCAATGTCGTGGACAGC	299-340	KF998171
Criv48*	$(AAG)_7$	F: CCACCAGCTCACTGCTTTAAC R: TCGGAATGTCGGTTTATTCGG	338-340	KF998174
Cama loci (*C. amara*-specific loci)				
Cama1	$(AT)_{10}$	F: GGAGATCTTCGCGGTTTGAC R: GCAGCGCGGAAGAGTTG	208-226	KF998151
Cama8	$(AT)_{10}$	F: ACACCGCATAGGAACGAGC R: GTTAGCACATTTGGACCTGGG	215-229	KF998153
Cama11	$(CCT)_5C(ATC)_{11}$	F: GCACTAGTATCAAATCCGAGAACC R: ACAGAGTGGAAGTCTCGTGG	248-270	KF998154
Cama13	$(AT)_{10}$	F: CTCCAGCTGCATGACCAAG R: GGTGGCTCCAACCTACCC	339-359	KF998155
Cama15	$(AAG)_7$	F: GGAAGTCAACCACACATGG R: GTTTGTTTGCCACCTTCGTC	250-273	KF998156
Cama16*	$(AT)_{11}$	F: TCGACGGCCTTGTTGAATAC R: GTAGAGAGACGCTGCTCCG	244-268	KF998157
Cama17	$(AG)_{13}$	F: GCTACTGTATTCCCAAGATTCCAC R: GTCCGCTAAGCACGTATGG	135-162	KF998158
Cama21	$(ATC)_7$	F: GGTTGCTCCACAGTCATGG R: CTGTGCGCAGAAACGTACC	312-315	KF998160
Cama27	$(AT)_{10}$	F: AACTAATGCGTCCGTCCAC R: TCAATCTTGTGCCTCTTATCTGG	357-390	KF998163
Cama32	$(CGT)_{4...}(AAT)_8$	F: CCGCCTTGCACGATTCTAAC R: CGGAGGAGCGTAGATGC	173-181	KF998164
Cama34	$(AC)_{12}(AT)_4$	F: AGAATTCACATGGACTTACGC R: CGTTGGTTGCTTAACTATCCTCG	276-297	KF998165

(continued)

Marker	Repeat motif	Primer sequence (5'→3')	Allele size range (bp)	GenBank accession no
C-all loci (non-specific loci)				
Cama7	$(AAG)_8$	F: GACGAAGTATGAACCTGAGCAC R: TGACTGGGAATTGACCTTCG	308-320	KF998152
Cama19	$(AAC)_{5...}(ATC)_7$	F: ATGGCGTTGAAGCCGTTG R: CCGTTGACCGTGCTTTAC	291-297	KF998159
Cama23	$(AAG)_8$	F: CAGCACAAGTACAGATGCG R: GCTTCGAAAGCGAAGATCGG	148-164	KF998161
Cama25	$(AAG)_7$	F: CAAGAAAGCGGATGCGGG R: AAACACGCCTCACAAAGTTC	144-150	KF998162

Table S2. Genetic diversity measures for the sampled *Cardamine* taxa and populations. Population abbreviations are presented in Appendix 1. 2*n*, ploidy level (more details in Appendix 1); N_{ind}, number of sampled individuals; N_{phe}, number of observed multilocus allele phenotypes; *A*, total number of alleles from all loci; *A'*, average number of alleles per locus; A_i', average number of alleles per locus in an individual; *P*(%), percentage of polymorphic loci; Shan. div., Shannon diversity index; H_o, proportion of observed (partial) heterozygotes, averaged over loci.
A. Data from ten *C. amara*-specific (Cama), and six *C. rivularis* auct.-specific (Criv) loci. **B.** Data from four shared (C-all) microsatellite loci.

A.

Species/Population	2*n*	N_{ind}	N_{phe}	*A*	*A'*	A_i'	*P*(%)	Shan. div.	H_o
triploid *C. ×insueta* – Urnerboden									
(Criv loci)		91	16	13	2.17	1.82	67	2.19	0.54
(Cama loci)		91	3	17	1.70	1.00	60	0.36	0.00
hypohexaploid *C. ×schulzii* – Urnerboden									
(Criv loci)		22	1	19	3.17	3.17	0	0.00	1.00
(Cama loci)		22	1	10	1.00	1.00	0	0.00	0.00
hypopentaploid *C. ×schulzii* – Urnerboden									
(Criv loci)		6	1	17	2.83	2.83	0	0.00	0.83
(Cama loci)		6	1	10	1.00	1.00	0	0.00	0.00
triploid hybrid - Lej da Champfèr									
(Criv loci)		15	2	6	1.00	1.08	20	0.67	0.08
(Cama loci)		15	2	17	1.70	1.60	10	0.64	0.50
***C. rivularis* auct. –** Urnerboden									
Urnerboden	2*x*	97	61	15	2.50	1.72	83	3.56	0.50
***C. pratensis* –** Urnerboden									
Urnerboden	4*x*	15	13	32	5.33	1.98	100	2.40	0.49
***C. rivularis* auct.** – outside Urnerboden									
Valbella	2*x*	6	6	20	3.33	1.80	83	1.79	0.58
Maloja	2*x*	6	6	16	2.67	1.67	83	1.79	0.53
Lej da Staz	2*x*	10	2	7	1.17	1.10	10	0.67	0.10
Lej da Champfèr	2*x*	15	9	17	2.83	1.70	100	1.97	0.52
Weinebene	2*x*	5	5	18	3.00	1.58	67	1.61	0.51
Koralpe	2*x*	9	9	23	3.83	1.70	100	2.20	0.48
Falkertsee	4*x*	6	6	20	3.33	2.25	83	1.79	0.67
Plannersee	4*x*	6	6	23	3.83	2.03	83	1.79	0.61
C. pratensis – outside Urnerboden									
Trieben	4*x*	6	6	36	6.00	2.81	100	1.79	0.76
Flattnitz	4*x*	6	6	26	4.33	2.08	83	1.79	0.69
Vordernberg	4*x*	6	6	26	4.33	2.00	83	1.79	0.53
Wegscheid	6*x*	6	6	31	5.17	2.93	100	1.79	0.80

(continued)

C. amara* subsp. *amara									
Urnerboden	2*x*	55	42	49	4.90	1.55	100	3.61	0.55
Ste-Croix	2*x*	6	2	15	1.50	1.42	10	0.45	0.42
Wehrenbach	2*x*	9	4	26	2.60	1.40	70	1.15	0.40
Železná	2*x*	6	6	27	2.70	1.45	80	1.79	0.45
Vrbice	2*x*	5	5	28	2.80	1.46	80	1.61	0.46
Hradišťany	2*x*	6	4	21	2.10	1.43	70	1.24	0.43
Višňová	2*x*	6	5	35	3.50	1.62	90	1.56	0.62
Telgárt	2*x*	6	5	25	2.50	1.50	80	1.56	0.50
Capatanii	2*x*	6	6	24	2.40	1.19	70	1.79	0.19
Osnabrück	2*x*	6	3	23	2.30	1.53	80	0.87	0.53
C. amara* subsp. *austriaca									
Lej da Champfèr	4*x*	9	2	18	1.80	1.98	10	0.53	0.75
Flüelapass	4*x*	6	4	22	2.20	1.70	70	1.33	0.60
Koralpe	4*x*	6	5	27	2.70	1.97	100	1.56	0.75

B.

Species/Pop.	2*n*	N_{ind}	N_{phe}	*A*	*A*’	A_i’	*P*(%)	Shan. div.	H_o
triploid *C.* ×*insueta* – Urnerboden									
		91	6	10	2.50	1.37	75	1.47	0.33
hypohexaploid *C.* ×*schulzii* – Urnerboden									
		22	1	9	2.25	2.25	0	0.00	0.50
hypopentaploid *C.* ×*schulzii* – Urnerboden									
		6	1	6	1.50	1.50	0	0.00	0.25
triploid hybrid - Lej da Champfèr									
		15	1	7	1.75	1.75	0	0.00	0.50
***C. rivularis* auct. –** Urnerboden									
Urnerboden	2*x*	96	8	13	3.25	1.36	75	1.19	0.34
***C. pratensis* –** Urnerboden									
Urnerboden	4*x*	15	13	21	5.25	2.17	100	2.49	0.70
***C. rivularis* auct.** – outside Urnerboden									
Valbella	2*x*	6	2	7	1.75	1.75	50	0.45	0.71
Maloja	2*x*	6	5	8	2.00	1.75	50	1.56	0.67
Lej da Staz	2*x*	10	2	7	1.75	1.68	25	0.61	0.68
Lej da Champfèr	2*x*	15	5	8	2.00	1.67	50	1.44	0.63
Weinebene	2*x*	5	4	11	2.75	1.65	100	1.33	0.65
Koralpe	2*x*	9	9	13	3.25	1.75	100	2.20	0.72
Falkertsee	4*x*	6	6	13	3.25	2.17	75	1.79	0.67
Plannersee	4*x*	6	6	16	4.00	2.25	100	1.79	0.92
C. pratensis – outside Urnerboden									
Trieben	4*x*	6	6	22	5.50	2.50	100	1.79	0.79
Flattnitz	4*x*	6	6	17	4.25	2.17	100	1.79	0.88
Vordernberg	4*x*	6	6	19	4.75	2.58	100	1.79	0.88
Wegscheid	6*x*	6	6	21	5.25	2.38	100	1.79	0.91
C. amara* subsp. *amara									
Urnerboden	2*x*	55	26	15	3.75	1.44	100	3.45	0.44
Ste-Croix	2*x*	6	1	6	1.50	1.50	0	0.00	0.50
Wehrenbach	2*x*	9	3	7	1.75	1.39	75	1.00	0.39
Železná	2*x*	6	4	9	2.25	1.46	75	1.33	0.46
Vrbice	2*x*	5	3	6	1.50	1.20	50	1.05	0.20
Hradišťany	2*x*	6	3	7	1.75	1.38	50	0.87	0.38
Višňová	2*x*	6	5	10	2.50	1.44	100	1.56	0.44
Telgárt	2*x*	6	5	10	2.50	1.71	100	1.56	0.67
Capatanii	2*x*	6	5	9	2.25	1.25	100	1.56	0.25
Osnabrück	2*x*	6	3	7	1.75	1.17	75	0.87	0.17
C. amara* subsp. *austriaca									
Lej da Champfèr	4*x*	9	2	4	1.00	1.72	25	0.35	0.72
Flüelapass	4*x*	6	4	8	2.00	1.71	75	1.33	0.54
Koralpe	4*x*	6	3	9	2.25	1.79	50	1.01	0.63

Table S3. Allele-sharing patterns at selected microsatellite loci. Entries (fields with "1"s) indicate the allele presence, those in bold with superscripts highlight alleles shared by the hybrids and some of the progenitor species: ^{A}, *C. amara*-specific alleles; $^{R/P}$, *C. rivularis* auct./*C. pratensis*-specific alleles (asterisks indicate specific alleles distinguishing between these two species at the Urneboden site); ^{n}, non-specific alleles, shared by both progenitors. *Rivularis* Urnerb., *C. rivularis* auct. from Urnerboden; *pratensis* Urnerb., *C. pratensis* from Urnerboden; *rivularis* other loc., *C. rivularis* auct. from other localities; *rivularis* other loc., *C. rivularis* auct. from other localities; *amara* Urnerb., *C. amara* from Urnerboden; *amara* other loc., *C. amara* from other localities; ×*insueta*, *C.* ×*insueta* from Urnerboden; 6*x* ×*schulzii*, hypohexaploid *C.* ×*schulzii* from Urnerboden; 5*x* × *schulzii*, hypopentaploid *C.* ×*schulzii* from Urnerboden; Champfèr h. (2*n*=24), triploid hybrid from Lej da Champfèr.

Locus Cama17 (Cama locus)

Alleles/Taxon	136	142	150	152	154	156	158	160	162
×*insueta* (2*n*=24)	**1^{A}**								**1^{A}**
6*x* ×*schulzii* (2*n*=46)									**1^{A}**
5*x* ×*schulzii* (2*n*=38)									**1^{A}**
Champfèr h. (2*n*=24)				**1^{A}**					
amara Urnerb.	**1^{A}**			**1^{A}**	1			1	**1^{A}**
amara other loc.	**1^{A}**	1	1	**1^{A}**		1	1	1	**1^{A}**

Locus Criv17 (Criv locus)

Alleles/Taxon	218	260	299	302	311	314	323	326	329	332	335	338	341
×*insueta* (2*n*=24)	**1$^{R/P}$**												
6*x* ×*schulzii* (2*n*=46)	**1$^{R/P}$**		**1$^{R/P*}$**										
5*x* ×*schulzii* (2*n*=38)	**1$^{R/P}$**												
Champfèr h. (2*n*=24)											**1$^{R/P}$**		**1$^{R/P}$**
rivularis Urnerb.	**1$^{R/P}$**												
pratensis Urnerb.	**1$^{R/P}$**		**1$^{R/P*}$**					1					
rivularis other loc.		1	**1$^{R/P}$**	1	1	1	1		1	1	**1$^{R/P}$**	1	**1$^{R/P}$**
pratensis other loc.			**1$^{R/P}$**	1									

Locus Cama25 (C-all locus)

Alleles/Taxon	145	148	151	154	157	160	163	166	169	172	175
×*insueta* (2*n*=24)	**1^{A}**		**1^{n}**				**1$^{R/P}$**	**1$^{R/P}$**			
6*x* ×*schulzii* (2*n*=46)	**1^{A}**			**1$^{R/P*}$**			**1$^{R/P}$**	**1$^{R/P}$**			
5*x* ×*schulzii* (2*n*=38)	**1^{A}**				**1$^{R*/P}$**			**1$^{R/P}$**			
Champfèr h. (2*n*=24)	**1^{A}**		**1^{n}**						**1$^{R/P}$**		
rivularis Urnerb.					**1$^{R*/P}$**	1	**1$^{R/P}$**	**1$^{R/P}$**	**1$^{R/P}$**		1
pratensis Urnerb.				**1$^{R/P*}$**		1	**1$^{R/P}$**	**1$^{R/P}$**	**1$^{R/P}$**	1	
rivularis other loc.		1	**1^{n}**	**1$^{R/P}$**		1	**1$^{R/P}$**	**1$^{R/P}$**	**1$^{R/P}$**		
pratensis other loc.		1		**1$^{R/P}$**		1	**1$^{R/P}$**	**1$^{R/P}$**	**1$^{R/P}$**	1	
amara Urnerb.	**1^{A}**	1	**1^{n}**								
amara other loc.	**1^{A}**	1	**1^{n}**								

Figure S1. Neighbour-joining tree of 266 individuals based on 10 *Cardamine amara*-specific (Cama) microsatellite loci and Bruvo distance. Colours indicate species/group affiliation. Ovals denoted STRU-1 - STRU-3 indicate genetic clusters as resolved by Bayesian clustering at optimal K=3. Adm. stands for genetic admixture (equivocal assignment).

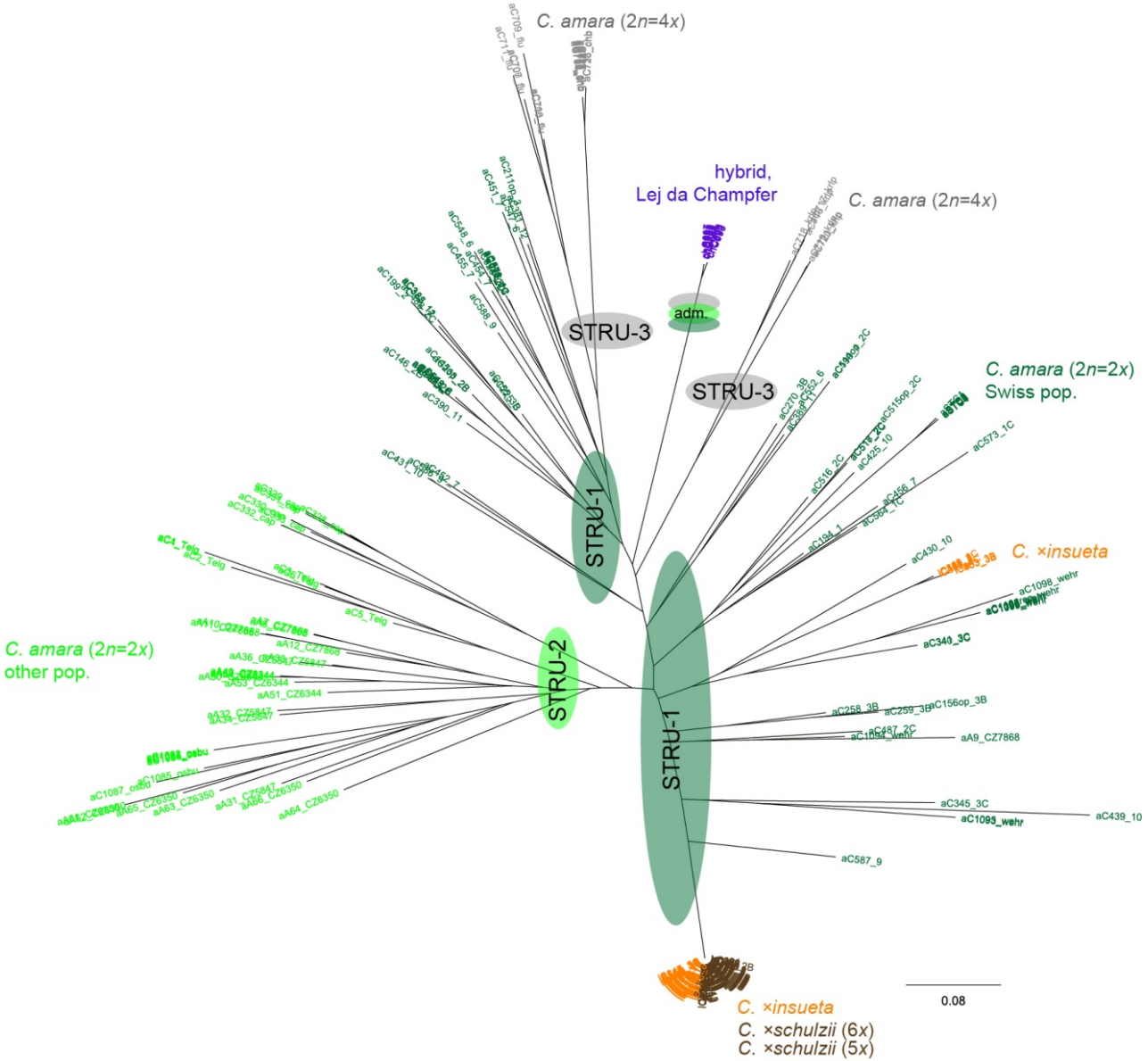

Figure S2. Principal coordinate analysis of 333 individuals based on six *Cardamine rivularis* auct./*C. pratensis*-specific (Criv) microsatellite loci and Jaccard coefficient. Percentage of variation represented by the axes is indicated. Colours and symbols indicate species/group affiliation.

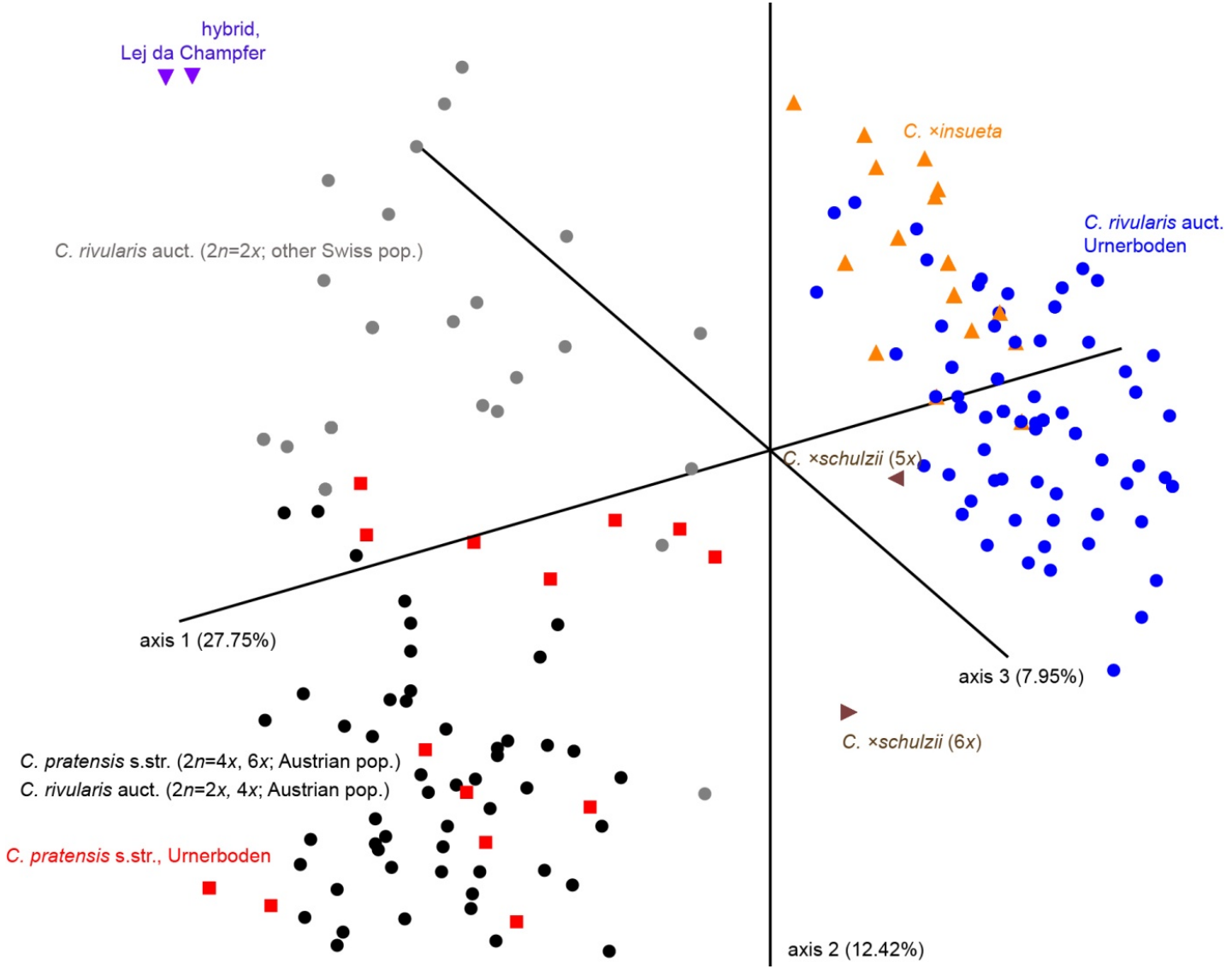

Figure S3. Neighbour-joining tree of 464 *Cardamine* individuals based on four shared (C-all) microsatellite loci and Jaccard coefficient. Colours indicate species/group affiliation. Coloured ovals denoted STRU-1 and STRU-2 indicate genetic clusters as resolved by Bayesian clustering at optimal K=2. Admixture stands for individuals showing genetic admixture.