Landoni, Beatrice and Viruel, Juan and Gomez, Rocio and Allaby, Robin G. and Brennan, Adrian C. and Pico, F. Xavier and Perez-Barrales, Rocio (2020) 'Microsatellite marker development in the crop wild relative Linum bienne using genome skimming.', Applications in plant sciences., 8 (5). e11349.

Further information on publisher’s website:
https://doi.org/10.1002/aps3.11349

Publisher’s copyright statement:
© 2020 Landoni et al. Applications in Plant Sciences is published by Wiley Periodicals, Inc. on behalf of the Botanical Society of America. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional information:

Use policy
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.
Microsatellite marker development in the crop wild relative Linum bienne using genome skimming

Beatrice Landoni1,6, Juan Viruel1, Rocio Gómez3, Robin G. Allaby4, Adrian C. Brennan5, Rocio Pérez-Barrales1, and F. Xavier Picó3

The genus Linum (Linaceae) includes 180–200 species, with most species diversity concentrated in the Mediterranean Basin. It has become an important plant group to investigate the evolution of breeding systems and genome duplication events (Sveinsson et al., 2014; Ruiz-Martín et al., 2018). Linum includes L. usitatissimum L., cultivated globally for fiber and oil, and its wild relative L. bienne Mill. (Fu, 2019). The two share a whole-genome duplication that occurred 5–9 mya (Sveinsson et al., 2014). Although phenotypic and genotypic variation of flax have been studied in relation to crop improvement (Fu, 2019), population variation in L. bienne remains relatively unexplored (but see Uysal et al., 2012).

Linum usitatissimum is an annual species, whereas L. bienne is a winter annual or perennial, growing in isolated patches across the Middle East, the Mediterranean Basin, and Western Europe (Uysal et al., 2012). For both species, seed production relies on self-pollination and, while outcrossing is rare, it has been central to the adaptation of the crop to northern latitudes by means of gene flow from L. bienne to L. usitatissimum (Gutaker et al., 2019). Sertse et al. (2019) highlighted the importance of eco-geographic factors in shaping L. usitatissimum genetic structure, and noted that the Mediterranean region is poorly represented in its core collection. Interestingly, the geographic distribution of L. bienne spans this area. Additionally, genotypic and phenotypic characterization of Turkish L. bienne populations has identified patterns of local adaptation (Uysal et al., 2010, 2012). Taken together, these studies reveal the potential value of L. bienne for crop improvement and evolutionary research. Most of the molecular tools available for L. bienne, including microsatellite markers, were retrieved ad hoc from those developed in L. usitatissimum, and population-level variation was not explored (Cloutier et al., 2012; Soto-Cerda et al., 2014). Only Uysal et al. (2010) genotyped L. bienne populations with inter-simple sequence repeat (ISSR) markers, but from a limited geographical range. Here, we screen 50 microsatellite markers that will serve to investigate genetic diversity and structure of L. bienne.

METHODS AND RESULTS

To identify microsatellite markers, we employed the approach used by Viruel et al. (2018), in which contigs are mined for microsatellite loci after a de novo assembly. DNA extractions for seven L. bienne individuals from different locations (Appendix 1) and corresponding whole genome shotgun libraries were prepared following the methods in Viruel et al. (2019). Equimolar sequencing read data was used for genome skimming, and 44 loci successfully amplified. Of these, 16 loci evenly spread across the L. usitatissimum reference nuclear genome were used for genotyping six L. bienne populations. Excluding one monomorphic locus, the number of alleles per locus ranged from two to 12. Four out of six populations harbored private alleles. The levels of expected and observed heterozygosity were 0.076 to 0.667 and 0.000 to 1.000, respectively. All 16 loci successfully cross-amplified in L. usitatissimum.

CONCLUSIONS: The 16 microsatellite loci developed here can be used for population genetic studies in L. bienne, and 28 additional loci that successfully amplified are available for further testing.

KEYWORDS: crop wild relative; Linaceae; Linum bienne; pale flax; population genetics; simple sequence repeat (SSR).
pooled libraries (150 × 150 bp) were sequenced at Novogene (Beijing, China) in an Illumina HiSeq X lane (Illumina, San Diego, California, USA). Contigs generated by assembling raw reads with SPAdes version 3.13 (Bankevich et al., 2012) were mapped against a *L. usitatissimum* nuclear genome reference (GenBank IDs CP027619.1–CP027633.1) in BWA version 0.7.17 (Li and Durbin, 2009). The mapping contigs were then scanned for di-, tri-, and tetranucleotide repeat motifs with MSATCOMMANDER version 1.0.8 (Faircloth, 2008) using default settings to design primers. Contigs containing microsatellite loci were filtered in R version 3.5.2 (R Core Team, 2018) using a custom-made script. Loci with primers that met the following requirements were retained: pair penalty <1.7, left-right penalty <0.8, difference in melting temperature <2°C, primer distance >20 bp, and pair product size between 89 and 301 bp. Polymorphic loci were then identified by BLASTing all contigs from locus >20 bp, and pair product size between 89 and 301 bp. The mapping contigs were then scanned for di-, tri-, and tetranucleotide repeat motifs with MSATCOMMANDER version 1.0.8 (Faircloth, 2008) using default settings to design primers. Contigs containing microsatellite loci were filtered in R version 3.5.2 (R Core Team, 2018) using a custom-made script. Loci with primers that met the following requirements were retained: pair penalty <1.7, left-right penalty <0.8, difference in melting temperature <2°C, primer distance >20 bp, and pair product size between 89 and 301 bp. Polymorphic loci were then identified by BLASTing all contigs mapping to the *L. bienne* reference genome for seven *L. bienne* individuals against the filtered contigs containing microsatellite loci, using BLAST version 2.2.31 (Altschul et al., 1990). Finally, 50 loci (Appendix 2) were left after filtering in R version 3.5.2 (R Core Team, 2018) using custom BLAST output. Only microsatellite loci with the following features were retained: ≥4 repeats of the base motif, <5 mismatches between BLAST match and reference, and at least one individual per BLAST group differed from the reference in number of motif repeats. The code used for de novo assembly and selection of microsatellite loci is available in Appendix S1.

For in vivo testing, DNA was extracted from seedlings of six *L. bienne* populations as well as other *Linum* species (Appendix 1). DNA extractions were performed with the ISOLATE II Plant DNA Kit (Bioline, London, United Kingdom), using approximately 20 mg of dry leaf material and following the kit protocol with buffer PAI. The 50 loci were first amplified in seven individuals following the Taq DNA Polymerase Master Mix instructions (ThermoFisher Scientific, Waltham, Massachusetts, USA). The PCR program consisted of an initial denaturation of 2 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at 56°C (annealing temperature [*T*$_a$]); and 2 min at 72°C; and a final extension step of 10 min at 72°C. For 12 out of 50 primer pairs, these conditions did not lead to amplification or produced multiple bands. When multiple bands were obtained, we tested the primers again by increasing *T*$_a$ by 1°C. In situations where no initial amplification occurred, we decreased *T*$_a$ by 1°C. In total, 44 loci amplified successfully at the end of this process (Appendix 2), with sizes as expected from MSATCOMMANDER.

### TABLE 1. Characteristics of 17 microsatellite loci developed for *Linum bienne* via genome skimming using the *L. usitatissimum* genome as a reference to identify a putative chromosome for each locus.

| Chromosome | Locus | Primer sequences (5’-3’) | Repeat motif | Allele size range (bp) | Mix | Fluorescent dye | GenBank accession no. |
|------------|-------|-------------------------|--------------|-----------------------|-----|----------------|----------------------|
| chr1       | ssr1.4 | F: CGAGCTGGGATCTTCCGAG  | (AGC)$_5$    | 127–136               | 4   | PET            | MN450483             |
|            |       | R: AGAACTGGAATGCGGCC    |              |                       |     |                |                      |
| chr2       | ssr2.1 | F: AAAGAAAGGAGGAGGAGG   | (AG)$_5$     | 215–233               | 1   | PET            | MN450485             |
|            |       | R: GGTCTAATTCGCTAGGGGC  |              |                       |     |                |                      |
| chr2       | ssr2a.2| F: CCGGTTGTTCCTTCAGCGCT| (AG)$_5$     | 280–282               | 2   | PET            | MN450486             |
|            |       | R: CACCTTGACGCCGAGCTCG  |              |                       |     |                |                      |
| chr2       | ssr2b.2| F: CCGGTTGTTCCTTCAGCGCT| (AG)$_5$     | 331–337               | 2   | PET            | MN450486             |
|            |       | R: CACCTTGACGCCGAGCTCG  |              |                       |     |                |                      |
| chr3       | ssr3.2 | F: GTCTGATAGCTGAGCACGAG | (AT)$_5$     | 153–163               | 2   | VIC            | MN450489             |
|            |       | R: GATAGGGTCTGGTTGAGGC  |              |                       |     |                |                      |
| chr3       | ssr3.4 | F: CGATCAACCAAGGCTGTTTCC| (AT)$_4$     | 226–252               | 4   | VIC            | MN450487             |
|            |       | R: ATGCTGTGTTCAAGACCAG  |              |                       |     |                |                      |
| chr4       | ssr4.2 | F: TCACAGCTTGGGCTGCTGG  | (AT)$_5$     | 200–206               | 2   | NED            | MN450493             |
|            |       | R: AGGCTGACGCTGCAAGGCC  |              |                       |     |                |                      |
| chr4       | ssr4.3 | F: ATAGTGACTGCAGTGTGACG| (AT)$_5$     | 127–130               | 3   | NED            | MN450492             |
|            |       | R: TCTGAAGACAGACCGTACTG |              |                       |     |                |                      |
| chr6       | ssr6.1 | F: TTACAGGGGATGTAACG    | (AAG)$_5$    | 157–163               | 1   | VIC            | MN450500             |
|            |       | R: ACTAGTGACTGCAGTGTGACG|              |                       |     |                |                      |
| chr9       | ssr9.3 | F: TACGCCAAGGCAGACAC    | (AC)$_4$     | 185–187               | 3   | VIC            | MN450514             |
|            |       | R: CACATACTAACTACCAACGC|              |                       |     |                |                      |
| chr10      | ssr10.1| F: TCTACAGGCGCTGACTGGG  | (AG)$_5$     | 119–127               | 1   | NED            | MN450518             |
|            |       | R: CGATCGGCTACGGGTATTG  |              |                       |     |                |                      |
| chr11      | ssr11.1| F: CTGATCCGCTGTGGAGG    | (AAC)$_5$    | 187–193               | 1   | FAM            | MN450519             |
|            |       | R: CATTGGCTGGAGCAGATGG  |              |                       |     |                |                      |
| chr11      | ssr11.2| F: TGCTGCAAAATGCTGTAGG  | (AAC)$_5$    | 243–264               | 2   | FAM            | MN450520             |
|            |       | R: ACCACATTCTTTCCACAC   |              |                       |     |                |                      |
| chr11      | ssr11.4| F: AAACCAACTCTCCACTTGGG | (AG)$_4$     | 292–298               | 4   | NED            | MN450521             |
|            |       | R: TCTCACTGAAAAAACCGCTTG|              |                       |     |                |                      |
| chr12      | ssr12.3| F: GGCACGAAATTTTTTCAGTC| (AAG)$_5$    | 219–225               | 3   | NED            | MN450523             |
|            |       | R: TGGAGAAGACAGATGCAGGC |              |                       |     |                |                      |
| chr12      | ssr12.4| F: CTACCTCGCTATCCGCAGTG| (AG)$_5$     | 174–194               | 4   | FAM            | MN450522             |
|            |       | R: TTGTCGCACCTCCTCAAGCC |              |                       |     |                |                      |
| chr14      | ssr14.3| F: ACATCCGGAACCTGATCCGG | (ACT)$_4$    | 280                    | 3   | FAM            | MN450527             |
|            |       | R: CGCTTATGTGGTGAAAGGG  |              |                       |     |                |                      |

aFor all primer pairs, the annealing temperature was 56°C.

bLoci were pooled into four groups (mixes 1 to 4) for capillary electrophoresis.

cFor each capillary electrophoresis mix containing four loci, four different dyes (PET, VIC, NED, FAM) were used to tag the reverse primer of each pair to facilitate genotyping.

dLocus 14.3 was monomorphic across all populations, so genetic diversity parameters were not computed for this locus.
output. To genotype all individuals, 16 loci were selected (Table 1) based on maximizing dispersion along the genome, the visual identification of polymorphisms on agarose gels, and avoiding the overlap of peaks during capillary electrophoresis by varying the PCR product sizes. PCR products were pooled in mixes of four loci, and reverse primers were tagged with four different fluorochromes (Table 1). PCR products were electrophoresed on an ABI PRISM 3700 DNA analyzer (Applied Biosystems, Foster City, California, USA), along with a GeneScan 500 LIZ fluorescent internal size standard. Transferability was also tested in three additional Linum species, including L. usitatissimum (Appendix 1), for the subset of 16 loci.

Genotyping was conducted manually in Peak Scanner Software version 1.0 (Applied Biosystems). Genetic diversity analyses are presented in Table 2. Allele number and observed heterozygosity (\( H_o \)) were estimated with the R package hierfstat version 0.04-22 (Goudet, 2005). Unbiased expected heterozygosity (\( H_s \)), departure from Hardy–Weinberg equilibrium (HWE), linkage disequilibrium, and number of private alleles were calculated using the R package poppr version 2.8.3 (Kamvar et al., 2014).

All 16 loci selected for genotyping amplified in L. bienne (1.23% of missing data on average), but cross-amplification was successful only in L. usitatissimum. In L. bienne, locus ssr14.3 was monomorphic and therefore excluded from the analyses. Locus ssr2.2 included two different microsatellite regions that were then treated as independent loci (ssr2a.2 and ssr2b.2). The number of alleles per locus varied between two and 12 over all six L. bienne populations. All populations harbored one to three private alleles for one or more loci, except for populations VII and IOW2. Depending on the population, 12 to 16 loci significantly deviated from HWE (\( P < 0.05 \)). When loci were in HWE, it was mostly due to fixed alleles (Table 2). \( H_s \), \( H_o \), and \( H_p \) ranged between 0.000 and 1.000, and \( H_s \) ranged between 0.000 and 0.773, across populations and loci. Linkage disequilibrium fluctuated between −0.336 and 1.000, with varying percentages of loci pairs in linkage disequilibrium within populations (between 9% and 54%, \( P < 0.05 \)) (Appendix S2).

The high \( H_s \) and frequent deviation from HWE (Table 2) might arise from fixed alleles on different paralogs produced by past polyploidization events in the genus Linum, which was also observed by Cloutier et al. (2012). If duplication is assumed when genotyping, consistency is essential while scoring loci showing a heterozygote fingerprint. Whether the latter is considered the result of homozgyosity, heterozygosity, or a combination of both at the duplicated locus will affect estimates of allele frequencies.

**CONCLUSIONS**

Microsatellite loci are ideal for providing fine-scale geographic and temporal information about population genetic processes such as relatedness. The set of loci developed here are distributed across the genome and will therefore be useful to distinguish between genome-wide processes caused by demography and locus-specific processes such as adaptation. However, putative paralogy needs investigation. The sequencing of different alleles and additional analysis of the genomic data set could serve to discriminate between paralog copies.
ACKNOWLEDGMENTS

The generation of genomic libraries used for genome skimming was possible thanks to a grant awarded by the Institute of Biological and Biomedical Science (IBBS, University of Portsmouth) to R.P.B and J.V. through the Research and Innovation Program. The Genetics Society provided support to B.L for bioinformatics training (Genetics Society Training Grant). The microsatellite lab work was funded by the Agencia Estatal de Investigación/Fondo Europeo de Desarrollo Regional (AEI/FEDER) of the European Union (grant CGL2016-77720-P to F.X.P), and by the Erasmus+ Placement program to B.L.

AUTHOR CONTRIBUTIONS

R.P.B., A.C.B., B.L., and R.G.A. collected the plant material; R.G., B.L., and J.V. conducted the lab work; B.L and J.V. implemented the genome skimming pipeline; B.L. analyzed the data and wrote the manuscript; R.P.B. and F.X.P. provided the funding and coordinated the work; all authors contributed to reviewing the manuscript.

DATA AVAILABILITY

Raw reads used for genome skimming were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID: PRJNA580472). Sequence information for microsatellite loci was deposited in NCBI’s GenBank, and accession numbers are provided in Table 1 and Appendix 2.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Script for de novo genome assembly using six Linum bienne individuals of different geographical origin, subsequent microsatellite loci mining and primer design, and in silico genotyping.

APPENDIX S2. Index of association for 16 polymorphic loci included in this study.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3): 403–410.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455–477.

Cloutier, S., R. Ragupathy, E. Miranda, N. Radovanovic, E. Reimer, A. Walchnowski, K. Ward, et al. 2012. Integrated consensus genetic and physical maps of flax (Linum usitatissimum L.). Theoretical and Applied Genetics 125: 1783–1795.

Faircloth, B. C. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8(1): 92–94.

Fu, Y.-B. 2019. A molecular view of flax gene pool, p. 270. In C. Cullis [ed.], Genetics and genomics of Linum, 1st ed. Springer Nature, Cham, Switzerland. Goudet, G. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes 2: 184–186.

Gutaker, R. M., M. Zaidem, Y. Fu, A. Diederichsen, O. Smith, R. Ware, and R. G. Allaby. 2019. Flax latitudinal adaptation at FlL11 altered architecture and promoted fiber production. Scientific Reports 9(1): 976.

Kamwar, Z. N., J. F. Tahima, and N. J. Grünwald. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: 281.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754–1760.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website https://www.R-project.org/.

Ruiz-Martín, J., R. Santos-Gally, M. Escudero, J. J. Mídgley, R. Pérez-Barrales, and J. Arroyo. 2018. Style polymorphism in Linum (Linaceae): A case of Mediterranean parallel evolution? Plant Biology 20(4): 100–111.

Sertse, D., F. M. You, S. Ravichandran, and S. Cloutier. 2019. The genetic structure of flax illustrates environmental and anthropogenic selections that gave rise to its eco-geographical adaptation. Molecular Phylogenetics and Evolution 137(4): 22–32.

Soto-Cerda, B. J., A. Diederichsen, S. Duguid, H. Booker, G. Rowland, and S. Cloutier. 2014. The potential of pale flax as a source of useful genetic variation for cultivated flax revealed through molecular diversity and association analyses. Molecular Breeding 34(4): 2091–2107.

Sveinsson, S., J. McDill, G. K. S. Wong, J. Li, X. Li, M. K. Deyholos, and Q. C. B. Cronk. 2014. Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (Linum) using transcriptomics. Annals of Botany 113(5): 753–761.

Uysal, H., Y. B. Fu, O. Kurt, G. W. Peterson, A. Diederichsen, and P. Kusters. 2010. Genetic diversity of cultivated flax (Linum usitatissimum L.) and its wild progenitor pale flax (Linum bienne Mill.) as revealed by ISSR markers. Genetic Resources and Crop Evolution 57(7): 1109–1119.

Uysal, H., O. Kurt, Y.-B. Fu, A. Diederichsen, and P. Kusters. 2012. Variation in phenotypic characters of pale flax (Linum bienne Mill.) from Turkey. Genetic Resources and Crop Evolution 59: 19–30.

Viruel, J., A. Hagenauver, M. Juin, F. Mirleau, D. Bouteiller, M. Boudagher-Kharrat, L. Ouahmane, et al. 2018. Advances in genotyping microsatellite markers through sequencing and consequences of scoring methods for Ceratonia siliqua (Leguminosae). Applications in Plant Sciences 6(12): e01201.

Viruel, J., M. Conejero, O. Hidalgo, L. Pokorny, R. F. Powell, F. Forest, M. B. Kantar, et al. 2019. A target capture-based method to estimate ploidy from herbarium specimens. Frontiers in Plant Sciences 10: 937.
## APPENDIX 1. Voucher information for populations of *Linum bienne*, *L. usitatissimum*, and other *Linum* species used in this study.

| Species          | Population* | n  | Locality                                      | Latitude    | Longitude   | Altitude (m) | Voucher accession no.b | Samplec |
|------------------|-------------|----|-----------------------------------------------|-------------|-------------|---------------|-------------------------|---------|
| *L. bienne* Mill.| 6           | 23 | Constantina-Cazalla de la Sierra, Seville, Spain | 37.93551111 | −5.711172222 | 529           | 202011                  | —       |
| *L. bienne*      | 11          | 23 | La Aliseda, Finca La Inmediata (Km 3), Jaén, Spain | 38.33105278 | −3.580855556 | 710           | 202012                  | L17     |
| *L. bienne* IOW2 | 24          |    | Bembridge, Isle of Wight, United Kingdom       | 50.68183333 | −1.074916667 | 9             | 202013                  | —       |
| *L. bienne*      | LLA         | 24 | Llanes, Asturias, Spain                        | 43.407375   | −4.687527778 | 26            | 202014                  | L58     |
| *L. bienne*      | SUT         | 30 | Sutton, Nottinghamshire, United Kingdom        | 53.35291111 | −0.959269444 | 15            | 202015                  | —       |
| *L. bienne* VIL  | 29          |    | Villeneuve, Charente Maritime, France           | 45.09393056 | −1.050338889 | 21            | 202016                  | L49     |
| *L. bienne*      | L01         | 1  | Pierrefeu-du-Ver, Provence-Alpes-Côte d’Azur, France | 43.25533    | 6.23802     | 200           | 202017                  | L01     |
| *L. bienne*      | CGA1        | 1  | Capo Gallo, Palermo, Sicily, Italy             | 38.2165     | 13.32183333 | 53            | 202018                  | L68     |
| *L. bienne*      | TYM         | 1  | Ty Mawr Holiday Park, Debinghshire, United Kingdom | 53.30307222 | −3.553280556 | 5             | 202019                  | L46     |
| *L. bienne* W77  | 1           |    | Greece                                        | 40.0875     | 21.722222   | 835           | Collection Gutaker et al. (2019) | L80     |
| *L. usitatissimum* | Cultivars Aramis and Volga | 2  | Terre de Lin, Saint-Pierre-Le-Viger, France    | 46.227638   | 2.213749    | 100           | 2020110                 | —       |
| *L. usitatissimum* | Cultivar Gisa and Primus | 2  | Italy                                         | 41.87194    | 12.56738    | —             | 260080 and 247707       | —       |
| *L. usitatissimum* | Cultivar Rabo189 | 1  | Morocco                                       | 31.791702   | −7.09262    | —             | 247713                  | —       |
| *L. suffruticosum* L. | —           | 6  | Puerto de las Palomas, Sierra de Grazalema, Cádiz, Spain | 36.80       | −5.41       | 400           | 1449143 and 1054224     | —       |
| *L. tenue* Desf.| —           | 9  | El Castillejo Botanical Garden, El Bosque, Cádiz, Spain | 36.765210   | −5.498114   | 298           | Live collection         | —       |

*Linum bienne* populations used for genotyping in vivo are in bold.

For populations 6, 11, IOW2, LLA, VIL, CGA1, L01, and TYM, vouchers were deposited in Portsmouth Natural History Museum (PORMG, Portsmouth, United Kingdom); for *L. usitatissimum*, the registered cultivars Aramis and Volga were provided by the cooperative Terre de Lin (Saint-Pierre-Le-Viger, France); the cultivars Gisa, Primus, and Rabo189 were provided by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany), for which herbarium sheets are available at the Genebank Information System of the institute and via the European Search Catalogue for Plant Genetic Resources (EURISCO); for *L. sulfurticosum*, a voucher is available at CSIC-Real Jardín Botánico (MA, Madrid, Spain); W77 and *L. tenue* are part of a private (Gutaker et al., 2019) and live (El Castillejo Botanical Garden) collections, respectively. For *L. usitatissimum* cultivars, coordinates reflect the centroids of the country of origin.

Populations used for genome skimming are marked with the name of the individual used. These names were also used to mark the contigs deposited in GenBank (Appendix 2). A dash means that the population was not used for genome skimming.
APPENDIX 2. Characteristics of 50 microsatellite loci tested in *Linum bienne*.

| Locus | Contig | Chromosome<sup>a</sup> | Repeat motif | Forward primer | Reverse primer | T<sub>i</sub> (°C) | Product size<sup>b</sup> | GenBank accession no. |
|-------|--------|-------------------------|--------------|----------------|----------------|----------------|----------------------|----------------------|
| chr1_L46_NODE_14803 | chr1 | AGG | ACTCTCATCCACATGACG | CGTGTGAACTGCTGACCAG | X | 201 | MN450480 |
| chr1_L46_NODE_28525 | chr1 | AGG | CTGACCTTCTCCGCAAGTACG | TGTGAACTGCTGACCAG | 56 | 240 | MN450481 |
| chr1_L68_NODE_33844 | chr1 | ATG | TTGAGGACTTTATGATGTTG | TTGAGGACTTTATGATGTTG | X | 231 | MN450482 |
| chr1_L80_NODE_42754 | chr1 | ACG | AGGAGCGTAAAATGCTG | AGGAGCGTAAAATGCTG | 56 | 228 | MN450484 |
| chr2_L46_NODE_99688 | chr2 | ATG | AGGGCAGTAAATGCTG | AGGGCAGTAAATGCTG | 155 | 228 | MN450488 |
| chr6_L46_NODE_157229 | chr6 | AGG | AGGCCTGAGATCTGACAAC | AGGCCTGAGATCTGACAAC | 162 | 186 | MN450498 |
| chr6_L46_NODE_95299 | chr6 | AGG | CCGTACACAACTGCTG | CCGTACACAACTGCTG | 244 | 244 | MN450499 |
| chr7_L46_NODE_17919 | chr7 | AGG | ACTCTCATCCACATGACG | AGGAGCGTAAAATGCTG | 267 | 207 | MN450496 |
| chr7_L46_NODE_21601 | chr7 | AGG | ACTCTCATCCACATGACG | AGGAGCGTAAAATGCTG | 160 | 244 | MN450497 |
| chr8_L46_NODE_40486 | chr8 | AGG | TTAAACTCTCTCTTCTCG | TTAAACTCTCTCTTCTCG | 254 | 254 | MN450501 |
| chr8_L46_NODE_63836 | chr8 | AGG | TTAAACTCTCTCTTCTCG | TTAAACTCTCTCTTCTCG | 254 | 254 | MN450501 |
| chr8_L46_NODE_7229 | chr8 | AGG | TTAAACTCTCTCTTCTCG | TTAAACTCTCTCTTCTCG | 162 | 162 | MN450501 |
| chr9_L46_NODE_67521 | chr9 | AGG | CGGCCTGCAACTGCTG | CGGCCTGCAACTGCTG | 199 | 199 | MN450507 |
| chr9_L46_NODE_95299 | chr9 | AGG | GCCGACAGAACATGCTG | GCCGACAGAACATGCTG | 244 | 244 | MN450507 |
| chr10_L100_NODE_15945 | chr10 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 250 | 250 | MN450517 |
| chr10_L100_NODE_14187 | chr10 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 184 | 184 | MN450524 |
| chr12_L64_NODE_36164 | chr12 | AGG | TCAACCTTCTCCGCACTTAC | TCAACCTTCTCCGCACTTAC | 210 | 210 | MN450510 |
| chr12_L46_NODE_17919 | chr12 | AGG | ACTCTCATCCACATGACG | AGGAGCGTAAAATGCTG | 297 | 297 | MN450511 |
| chr12_L46_NODE_21601 | chr12 | AGG | ACTCTCATCCACATGACG | AGGAGCGTAAAATGCTG | 256 | 256 | MN450512 |
| chr12_L58_NODE_35466 | chr12 | AGG | ATCTACGACGATCTGACAAG | ATCTACGACGATCTGACAAG | 168 | 168 | MN450513 |
| chr12_L58_NODE_25256 | chr12 | AGG | ATCTACGACGATCTGACAAG | ATCTACGACGATCTGACAAG | 166 | 166 | MN450515 |
| chr13_L68_NODE_95299 | chr13 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 183 | 183 | MN450516 |
| chr13_L80_NODE_109677 | chr13 | AGG | TCAACCTTCTCCGCACTTAC | TCAACCTTCTCCGCACTTAC | 155 | 155 | MN450518 |
| chr14_L100_NODE_48466 | chr14 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 178 | 178 | MN450528 |
| chr15_L100_NODE_40627 | chr15 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 163 | 163 | MN450529 |
| ssr1_L68_NODE_46821 | chr16 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 129 | 129 | MN450483 |
| ssr10_L100_NODE_100690 | chr16 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 120 | 120 | MN450518 |
| ssr11_L146_NODE_43040 | chr16 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 155 | 155 | MN450519 |
| ssr12_L114_NODE_100592 | chr16 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 243 | 243 | MN450520 |
| ssr14_L114_NODE_14339 | chr16 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 300 | 300 | MN450521 |
| ssr12_L146_NODE_28661 | chr17 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 225 | 225 | MN450523 |
| ssr12_L146_NODE_38654 | chr17 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 173 | 173 | MN450522 |
| ssr14_L100_NODE_12417 | chr17 | AGG | ACCACCAAGCTGCACTG | ACCACCAAGCTGCACTG | 279 | 279 | MN450527 |
| ssr21_L46_NODE_13038 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 215 | 215 | MN450548 |
| ssr22_L48_NODE_29522 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 277 | 277 | MN450486 |
| ssr3_L68_NODE_6280 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 157 | 157 | MN450489 |
| ssr5_L64_NODE_32336 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 227 | 227 | MN450487 |
| ssr24_L49_NODE_25476 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 198 | 198 | MN450492 |
| ssr34_L49_NODE_22236 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 161 | 161 | MN450500 |
| ssr56_L64_NODE_173634 | chr17 | AGG | TCTAGACGCTGCACTG | TCTAGACGCTGCACTG | 185 | 185 | MN450514 |

<sup>a</sup>Product size reported here are based on MSATCOMMANDER output, although the sizes were double-checked by looking at the agarose gels of the PCR products for all loci, where a ladder was added to assist the estimation of the products’ approximate size.

<sup>b</sup>The loci were obtained via genome skimming using the L. *lotus* genome as reference; therefore, it was possible to identify a putative chromosome for each locus.

<sup>c</sup>The product sizes reported here are based on MSATCOMMANDER output, although the sizes were double-checked by looking at the agarose gels of the PCR products for all loci, where a ladder was added to assist the estimation of the products’ approximate size.

Note: T<sub>i</sub> is optimized annealing temperature for each primer pair; X = unsuccessful amplification.