BMC Bioinformatics

SOFTWARE

Open Access

X-cross/over: a web tool for graph-based estimation of meiotic crossover events in plants

Szabolcs Makai[1], Diána Makai[1], Erika Chonata-Jiménez[1], Ildikó Karsai[1], Péter Mikó[1], Adél Sepsi[1*] and András Cseh[1*]

*Correspondence:
Adél Sepsi
sepsi.adel@atk.hun-ren.hu
András Cseh
cseh.andras@atk.hun-ren.hu
[1]HUN-REN Centre for Agricultural Research, Agricultural Institute, Martonvásár 2462, Hungary

Abstract

Background Crossovers are essential for genome stability and genetic diversity, yet in plants they occur infrequently, typically restricted to only one to three per chromosome pair. Genotyping approaches such as SNP arrays or genotyping-by-sequencing (GBS) enable high-resolution detection of crossover frequency, a critical step for elucidating the mechanisms that regulate meiotic recombination and for exploiting it in plant breeding. Despite their widespread use and the availability of highly reproducible marker sets, user-friendly tools for reliable recombination analysis remain scarce.

Results Here we present X-cross/over, a web-based platform that applies a graph-theoretical algorithm to estimate crossover frequencies from SNP datasets in HapMap format. The platform was evaluated using publicly available barley backcross inbred populations and newly developed wheat doubled haploid lines. Across both datasets, X-cross/over detected crossover events with high accuracy and sensitivity, yielding results consistent with published genotyping and cytological analyses. Importantly, the tool produces outcomes comparable to expert analyses while remaining accessible to users without bioinformatics expertise.

Conclusions X-cross/over provides a consistent and transparent framework for detecting crossover sites and quantifying their frequency. Implemented in a platform-independent environment, the application is freely available at https://insilicolabdesk.atk.kinin.hu, making it a versatile resource for exploring the genetic and epigenetic regulation of meiotic recombination across plant species.

Keywords Crossover, Crop, SNP array, GBS, Meiosis, DSB

Introduction

Meiosis is the specialized cell division responsible for ensuring genetic diversity in subsequent generations, achieved through the reshuffling of parental alleles during recombination. In this process, crossovers (COs) form through the repair of programmed DNA double-strand breaks (DSBs), ensuring faithful homolog segregation and generating new allele combinations [11, 12, 17]. In plants, crossover frequency and patterning are highly regulated, with typically one to three events per chromosome pair. Crossover positioning is strongly influenced by sequence polymorphism and local chromatin

states, resulting in their enrichment in gene-rich subtelomeric regions and suppression near centromeres and pericentromeres [4, 19]. Analysis of CO distribution is essential to uncover regulatory mechanisms and intraspecific variation in crossover numbers both to be exploited in modern breeding efforts for climate resilience.

High-throughput genotyping technologies allow fine-scale analysis of CO events using segregating markers in mapping populations such as recombinant inbred lines (RILs), backcross inbred lines (BILs), and doubled haploid (DH) lines, as well as through single-cell genotyping of F_1 pollen [10, 14, 15] (Fig. 1). SNP arrays provide reproducible, standardized marker sets, while genotyping-by-sequencing (GBS) offers a flexible and cost-effective approach for generating genome-wide SNP datasets [8]. Studies in *Arabidopsis* have demonstrated that COs preferentially accumulate near gene promoters, indicating that SNP arrays utilizing markers predominantly located within coding regions are effective to detect COs in plants [3]. Together, these platforms have supported the development of dense SNP maps across crop species, enabling studies of genetic diversity, population structure, and trait associations [1, 5, 7].

Despite the availability of large genomic datasets, accessible computational tools for detecting and quantifying COs remain limited. Available methods are often technically demanding, lack transparency, or are poorly adapted for plant datasets. Current CO detection commonly relies on multi-step pipelines that combine several genetic mapping tools and/or approaches in which crossovers are inferred from changes in the parental allelic state along with the physical length of the chromosome [2, 6, 10, 16, 18] rather than on a single-step, user-friendly solution. This creates a gap between data generation and its effective use for recombination studies and breeding applications.

To overcome this limitation, we developed X-cross/over, a platform-independent online tool that applies a graph-theoretical framework to detect and quantify meiotic COs from SNP array– and GBS-derived datasets. Designed to combine accuracy with usability, X-cross/over enables transparent crossover analysis without requiring

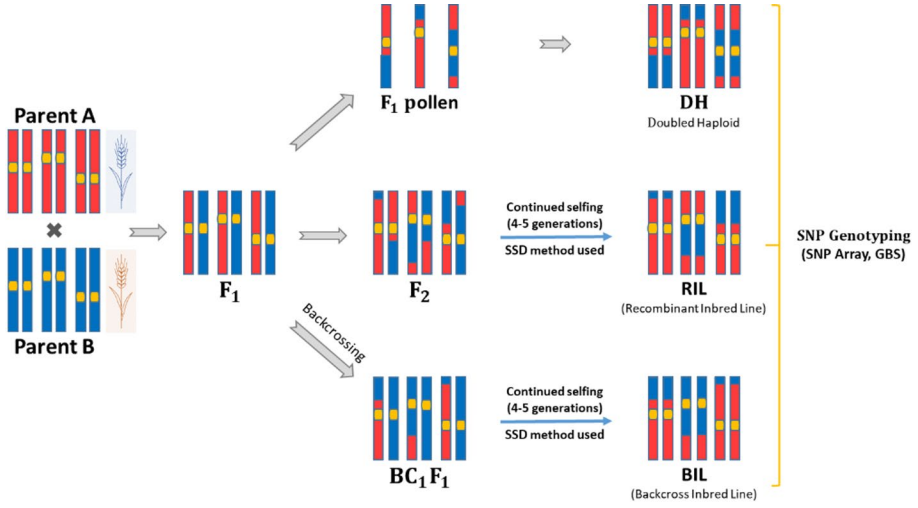

Fig. 1 Development of mapping populations required for the detection of meiotic crossovers (COs) using the X-cross/over tool. Single-cell genotyping of F_1 pollen provides a direct approach for detecting CO events. Alternatively, F_1 plants can be utilised to develop segregating populations. Doubled haploid (DH) lines can be obtained from F_1 pollen, while selfing F_2 progeny for 4–5 generations by single seed descent (SSD) yields recombinant inbred lines (RILs). Backcrossing the F_1 to a parent, followed by several generations of selfing, produces backcross inbred lines (BILs). All populations types can be genotyped with SNP arrays or genotyping-by-sequencing (GBS) platforms to infer crossover patterns

bioinformatics expertise, thereby extending the utility of genomic data for both fundamental research and applied breeding.

Implementation

X-cross/over is implemented as a client-side web application for analysing genome-wide crossover (CO) patterns from plant SNP datasets. All computations are executed locally within modern web browsers (e.g., Firefox, Edge, Chrome), avoiding server-side processing and thereby eliminating delays associated with file transfers.

The graphical user interface (GUI) provides two input fields for uploading SNP datasets: one for the parental lines and another for the derived progeny lines. Users can process datasets, reset inputs, and download results directly through the interface. An optional function can be activated to determine crossover positions by reporting the nearest flanking SNPs, and a dedicated message area provides progress updates and notifications.

Inputs and input validation

The software is designed for datasets in which crossovers can be traced unambiguously, including single-cell genotyping of F_1 pollen as well as segregating homozygous populations such as DHs, RILs, and BILs (Fig. 1).

SNP datasets generated by SNP arrays or genotyping-by-sequencing are supported in HapMap format [9]. Two input files are required: one containing parental genotypes and the other containing progeny genotypes. A template file is provided for convenience on the webpage.

Filtering polymorphic markers

The algorithm identifies polymorphic markers between parents and retains those meeting quality thresholds (polymorphic in ≥ 1 line and < 10% missing data). Each marker is assigned an ancestry value of −1 or 1 depending on parental match, with ambiguous cases receiving 0.

Scoring polymorphic markers

Polymorphic markers are scored based on their ancestry and that of neighbouring markers, using a dynamic window of 21 markers (10 upstream and 10 downstream) to increase the accuracy and avoid genotyping errors. Scores approach − 21 or 21 within parental haplotype blocks, while ambiguous markers contribute zero. For each marker m_i, ancestry values are summed over a window of 21 markers (10 upstream and 10 downstream).

$$score_i = \sum_{x=\max\,(1,i-k)}^{\min\,(n,i+k)} ancsestry\,(m_x)$$

where n is the total number of polymorphic markers on the chromosome, k = 10 = window half-size, the functions max and min ensure that the summation window does not extend beyond the first (1) or last (n) marker.

Each marker is also assigned a "range" value reflecting the genomic distance (in base pairs) spanned by its window of neighbouring markers. Specifically, for marker m_i with genomic coordinate p_i, the range is

$$range_i = position\left(m_{\min\,(n,\,i+k)}\right) - position\left(m_{\max\,(1,\,i-k)}\right)$$

Estimating crossovers

Scores are integrated into a chromosome-specific adjacency matrix, which encodes relationships among polymorphic markers as a graph. Breakpoints in marker ancestry profiles define crossover sites, localized to intervals flanked by the nearest SNPs.

For each sample *s* and chromosome *chr*, let

$$\mathrm{V} \;=\; \{m_1, \ldots, m_n\}$$

be the set of polymorphic markers ordered by genomic position. An $n \times n$ binary adjacency matrix A is constructed to define the graph

$$G_{s,chr} = \;(V,\, E)\,.$$

An edge $\{m_i, m_j\} \in E$ ($A_{ij} = 1$) is added when the following conditions hold:

1. Strong ancestry signal.

$$|\mathrm{score}\,(m_i)| \;>\; 8 \;\mid \mathrm{and} \mid \mathrm{score}\,(m_j)\,| > \;8|$$

   Each marker's window reflects a ≥ 70/30 ancestry imbalance. For example, if a marker has 4 non-similar neighbours, it must have at least 4 + 8 = 12 similar ones. With no ambiguous neighbours, the threshold is 14/20 = 0.7.

2. Comparable genomic scale.

$$|\mathrm{position}\,(m_i) - \mathrm{position}\,(m_j)| \leq \;\max\!\Big(\mathrm{range}\,(m_i)\,,\;\mathrm{range}\,(m_j)\Big).$$

   This limits edges to marker pairs separated by at most the larger of their effective window spans, analogous to a window-size filter.

3. Consistent ancestry sign.
   score (m_i) · score (m_j) > 0, ensuring both markers indicate the same parental origin.

$$\{m_i, m_j\} = \begin{cases} & |score\;(m_i)| > 8 \text{ and} \\ & |score\;(m_j)| > 8 \text{ and} \\ 1 \quad \text{if} & \\ 0 \quad \text{else} & |\mathrm{position}\,(m_i) - \mathrm{position}\,(m_j)| \leq \;\max\!\Big(\mathrm{range}\,(m_i)\,,\;\mathrm{range}\,(m_j)\Big) \text{and} \\ & (score\;(m_i) \times score\;(m_j)) > 0 \end{cases}$$

The resulting adjacency matrix typically contains clusters ("islands") of ones, which correspond to subgraphs (Fig. 2). We retain only subgraphs with at least four vertices. Each retained subgraph *Sg* is summarized by its leftmost (g_{leftg}) and rightmost (g_{right}) markers. After ordering markers within *Sg* by genomic position (*g1,…,gk*):

- If score(g_n) · score(g_{n+1}) < 0, then (g_n, g_{n+1}) flank a putative crossover site.

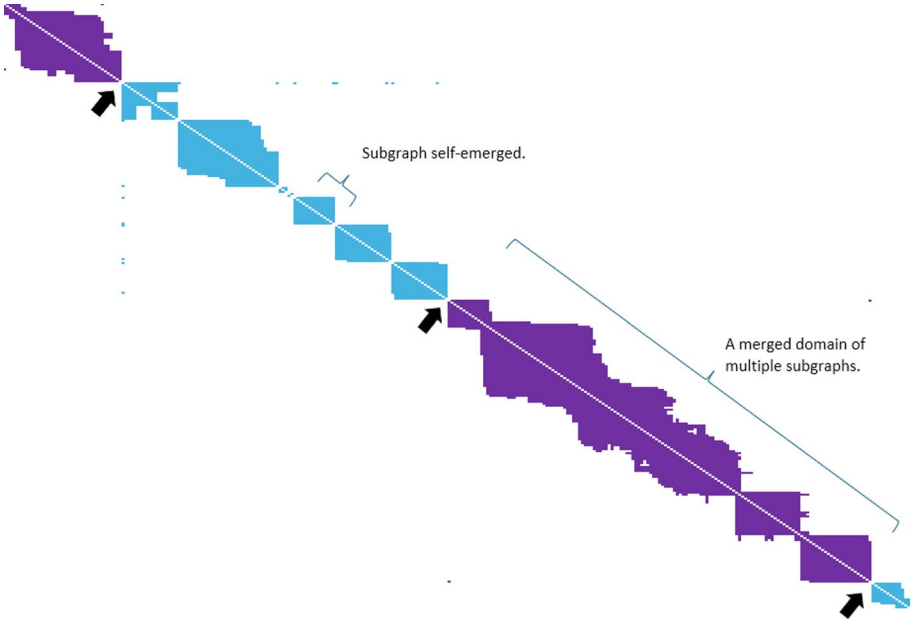

Fig. 2 Adjacency matrix of polymorphic markers on wheat chromosome 2A for a representative DH line (A173). Both axes correspond to the sequential order of polymorphic markers; Empty areas show zero values, while coloured areas indicate a value of one. Colours reflect the sign of the summed scores, illustrating parental dominance across chromosome segments. Four subgraphs are visible, flanking three crossovers (breakpoints), indicated by arrows. Subgraphs appear as distinct islands. They are merged when derived from the same parental background, revealing the overall crossover landscape

- If all adjacent products are positive, the region is treated as a single ancestry block and neighbouring subgraphs are merged (collapsed).

$G_{s,chr} = (V, E)$ s denotes the sample and chr the chromosome

V : the polymorphic markers (m) of the chromosome

$E \subseteq \{\{m_i, m_j\} | m_i, m_j \in V \text{and } i \neq j\}$, where

Outputs

The application produces summary and detailed reports. The basic report lists crossover counts by genotype and chromosome, while the detailed report provides all crossover site records, including flanking SNP IDs. The interface additionally displays marker statistics, progress updates, and error notifications. The analysis time depends on the size of the input HAPMAP files and the user's computer configuration, but it generally requires approximately 1–3 min.

Results

The tool was validated through direct in vivo comparison of the estimates and by reanalysing publicly available crossover datasets. We used the 50 K iSelect SNP data from barley BIL populations published by Dreissig et al. [6], available from the e!DAL repository (https://doi.org/10.5447/ipk/2019/20), together with the corresponding crossover datas et (Dataset S1, [6], Wiley Online Library Supplementary Material). Three populations -HEB1 (54 lines), HEB2 (45 lines), and HEB3 (55 lines)- were selected for validation. The SNP data sets were reformatted into HAPMAP files, and processed with X-cross/over to

estimate crossover numbers. Across all three population, a total of 32 250 SNP markers were submitted, of which 12 900 (HEB1), 15 374 (HEB2), and 15 727 (HEB3) were polymorphic, and used for downstream crossover analysis.

Bland–Altman analysis showed that X-cross/over estimates matched the previously published values at the chromosomal level in approximately 66% of cases across the three HEB populations (Fig. 3). The majority of points fell within the relatively narrow limits of agreement (−5.33 to + 3.76 COs), and no systematic trend was observed across the range of CO values, demonstrating that X-cross/over provides stable and consistent estimates. On average, estimates were only slightly lower than the published values (mean bias = − 0.78 COs), with the mean difference line lying just below zero, and the minor tendency towards underestimation was mostly associated with outliers reported in the reference dataset (Supplementary Table 1).

Line-level variance was consistently lower when using X-cross/over compared to the published method (HEB1: 18.21 vs. 82.25; HEB2: 22.87 vs. 63.48; HEB3: 15.02 vs. 64.08), resulting in tighter confidence intervals across all populations (Supplementary Table 1). The maximum CO numbers per line reported in the published dataset (56, 44, and 47 for HEB1—HEB3) were substantially reduced with X-cross/over (26, 32, and 25, respectively). These patterns were consistent across all three populations, supporting the conclusion that X-cross/over provides stable estimates of CO numbers at the chromosomal level.

To further validate the tool, we analysed our own wheat doubled haploid population genotyped with the 25 K iSelect SNP array, with raw SNP genotyping data deposited in the HUN-REN Data Repository Platform'ARP' (https://hdl.handle.net/21.15109/ARP

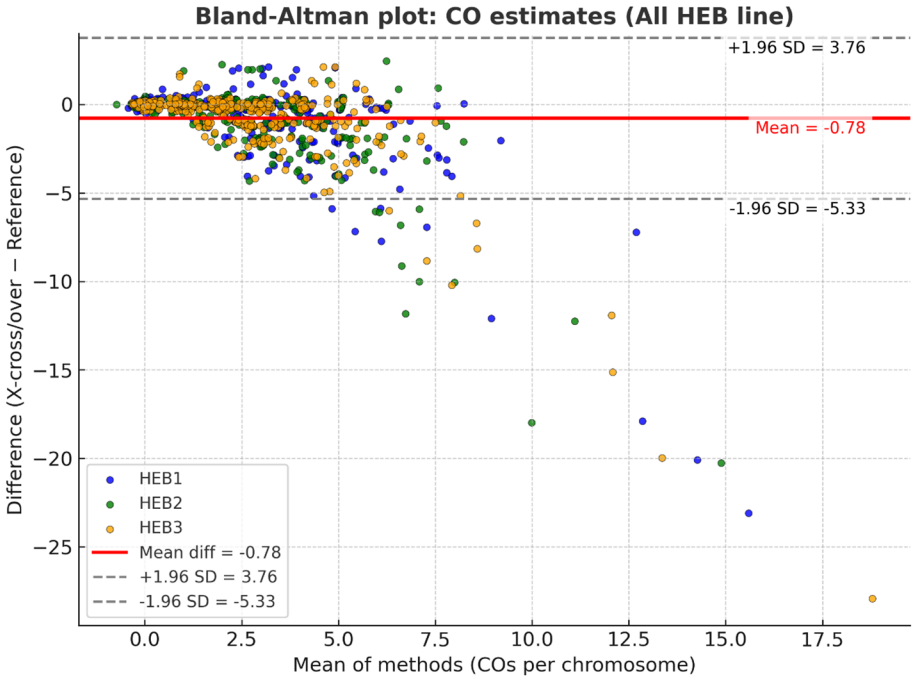

Fig. 3 Bland–Altman plot comparing crossover (CO) estimates obtained with the X-cross/over tool and previously published data across three barley BIL populations (HEB1–HEB3). The x-axis shows the mean number of COs per chromosome estimated by both methods, while the y-axis represents their difference (X-cross/over – Reference). Each point corresponds to one chromosome within an individual line; HEB1 (blue), HEB2 (green), and HEB3 (orange) are distinguished by color. The red line marks the mean difference, and grey dashed lines indicate the 95% limits of agreement (± 1.96 SD)

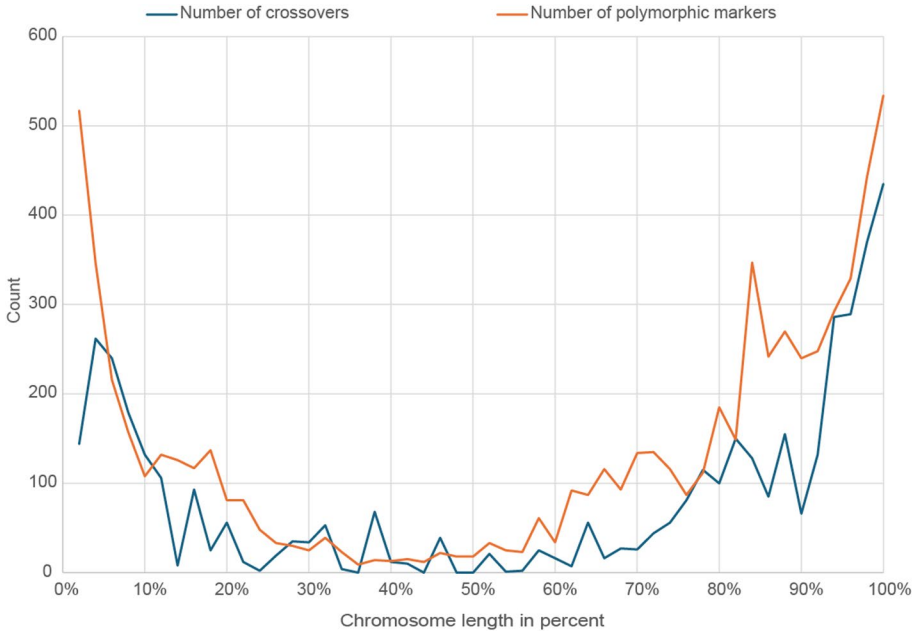

Fig. 4 Marker density and crossover (CO) events in 111 wheat doubled haploid lines. Marker density shows that crossover estimates reflect true recombination events rather than marker-density bias, with two illustrative examples in the 20–30% and 60–80% chromosomal-length regions

/8KZGSZ). A total of 21 695 markers were submitted, of which 6 986 were polymorphic across the 21 wheat chromosomes, representing 111 DH lines. These polymorphic markers served as the basis for subsequent crossover identification using the X-cross/over tool. Despite the lower marker density compared with the barley dataset, the tool generated reliable estimates with low variance (25.66), without adjustment of the default parameters (Supplementary Table 2). Importantly, crossover detection was not strictly dependent on local marker density: intervals with relatively high numbers of polymorphic markers did not always coincide with those showing a relatively high CO frequency (Fig. 4). This independence underscores the robustness of the X-cross/over for unbiased CO detection.

The average number of COs in the wheat DH population was 23.51 ± 5.09 per haploid genome, corresponding to 47.02 ± 10.18 per meiosis after adjustment. This estimate is consistent with sequence-based genotyping of 2 100 recombinant inbred lines (48.1 ± 0.18; [10]) and with cytological estimates based on of HEI10 foci counts in wheat meiocytes at pachytene (45.4 ± 7.1; [13]). Since HEI10 immunolocalization is a robust cytological marker for crossover designation, this strong concordance further validates the accuracy of the X-cross/over approach.

Conclusions

In summary, X-cross/over provides a consistent framework for estimating meiotic crossovers, as demonstrated by its reduced variance and the absence of extreme or population-specific discrepancies when compared with previously published data in barley and with crossover counts from our own wheat doubled haploid population. Importantly, the estimates obtained for wheat were in close agreement with cytological measurements of HEI10 foci, supporting the biological validity of the approach. The graph-based strategy which integrates ancestry-derived scores and neighbourhood criteria, enables reliable

detection and merging of consecutive ancestry blocks to identify crossover intervals. Optimized for widely used SNP array and genotyping-by-sequencing (GBS) platforms, X-cross/over is readily applicable to other crop species. While its performance on extremely sparse or highly heterogenous datasets remains to be tested, our results indicate that it can provide reproducible, high-resolution crossover landscapes to support both fundamental research and breeding applications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-025-06334-7.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
This work was supported and managed by HUN-REN Agricultural Research Centre, Agricultural Institute.

Author contributions
Conceptualization, AC, AS, SM; methodology, SM, DM, ECJ, IK, PM, AS; validation, SM, AS and AC; investigation, SM, DM, ECJ, IK, PM, AS, AC; data curation, SM, AC and AS; writing, SM, AC and AS; supervision, AC and AS.; funding acquisition, IK, PM and AC. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by the National Research, Development and Innovation Fund and Hungarian National Laboratory Program, Agribiotechnology project (grant number RRF-2.3.1-21-2022-00007). PM and AC were supported by János Bolyai Research Scholarships of the Hungarian Academy of Sciences (BO/00206/24/4 and BO/00416/23/4, respectively).

Data availability
The barley genotype dataset has been deposited in e!DAL under https://doi.org/10.5447/ipk/2019/20 [6]. The barley crossover dataset is available as Dataset S1 in Dreissig et al. [6]: https://nph.onlinelibrary.wiley.com/action/downloadSupplement%3Fdoi=10.1111%2Fnph.16810%26file=nph16810-sup-0001-DatasetsS1-S3.xlsx. The raw SNP genotyping data of the wheat DH lines generated in this study have been deposited in the HUN-REN Data Repository Platform (ARP) under handle: https://hdl.handle.net/21.15109/ARP/8KZGSZ. Availability and requirements: Project name: X-cross/over. Project home page: https://insilicolabdesk.atk.kinin.hu/. Operating system(s): Platform independent (in browser). Programming language: nodejs. Other requirements: NA. License: CC NC SA. Any restrictions to use by non-academics: no restriction. The data supporting the findings of this study are available within the manuscript or supplementary files, as well as from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 19 September 2025 / Accepted: 17 November 2025
Published online: 08 December 2025

References

1. Bayer MM, Rapazote-Flores P, Ganal M, Hedley PE, Macaulay M, Plieske J, et al. Development and evaluation of a barley 50k iSelect SNP array. Front Plant Sci. 2017;8:1792. https://doi.org/10.3389/fpls.2017.01792.
2. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19(7):889–90. https://doi.org/10.1093/bioinformatics/btg112.
3. Byun D, Son N, Kim H, Kim J, Park J, Park SJ, et al. COmapper: high-resolution mapping of meiotic crossovers by long-read sequencing in Arabidopsis. New Phytol. 2025;247(4):1942–57. https://doi.org/10.1111/nph.70304.
4. Choi K, Henderson IR. Meiotic recombination hotspots - a comparative view. Plant J. 2015;83(1):52–61. https://doi.org/10.1111/tpj.12870.
5. Cui F, Zhang N, Fan XL, Zhang W, Zhao CH, Yang LJ, et al. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. Sci Rep. 2017;7(1):3788. https://doi.org/10.1038/s41598-017-04028-6.

6. Dreissig S, Maurer A, Sharma R, Milne L, Flavell AJ, Schmutzer T, et al. Natural variation in meiotic recombination rate shapes introgression patterns in intraspecific hybrids between wild and domesticated barley. New Phytol. 2020;228(6):1852–63. https://doi.org/10.1111/nph.16810.
7. Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, et al. Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genom. 2011;12:302. https://doi.org/10.1186/1471-2164-12-302.
8. He J, Zhao X, Laroche A, Lu ZX, Liu, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci. 2014;5:484. https://doi.org/10.3389/fpls.2014.00484.
9. International HapMap Consortium. The International HapMap Project. Nature. 2003;426(6968):789–96. https://doi.org/10.1038/nature02168.
10. Jordan KW, Wang S, He F, Chao S, Lun Y, Paux E, et al. The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. Plant J. 2018;95(6):1039–54. https://doi.org/10.1111/tpj.14009.
11. Lambing C, Franklin FC, Wang CR. Understanding and manipulating meiotic recombination in plants. Plant Physiol. 2017;173(3):1530–42. https://doi.org/10.1104/pp.16.01530.
12. Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. The molecular biology of meiosis in plants. Annu Rev Plant Biol. 2015;66:297–327. https://doi.org/10.1146/annurev-arplant-050213-035923.
13. Osman K, Desjardins SD, Simmonds J, Burridge AJ, Kanyuka K, Henderson IR, et al. FIGL1 prevents aberrant chromosome associations and fragmentation and limits crossovers in polyploid wheat meiosis. New Phytol. 2024;244(2):528–41. https://doi.org/10.1111/nph.19716.
14. Pan Q, Ali F, Yang X, Li J, Yan J. Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS ONE. 2012;7(12):e52777. https://doi.org/10.1371/journal.pone.0052777.
15. Ren W, Gong X, Li K, Zhang H, Chen F, Pan Q. Recombination pattern characterization via simulation using different maize populations. Int J Mol Sci. 2020;21(6):2222. https://doi.org/10.3390/ijms21062222.
16. Taylor J, Butler D. R package ASMap: efficient genetic linkage map construction and diagnosis. J Stat Softw. 2017;79(6):1–29. https://doi.org/10.18637/jss.v079.i06.
17. Wang Y, Copenhaver GP. Meiotic recombination: mixing it up in plants. Annu Rev Plant Biol. 2018;69:577–609. https://doi.org/10.1146/annurev-arplant-042817-040431.
18. Waesch C, Gaede N, Gao Y, Ehle M, Himmelbach A, Fuchs J, et al. Population-wide single-pollen nuclei genotyping in rye sheds light on the genetic basis and environmental plasticity of meiotic recombination. New Phytol. 2025. https://doi.org/10.1111/nph.70656.
19. Ziolkowski PA, Henderson IR. Interconnections between meiotic recombination and sequence polymorphism in plant genomes. New Phytol. 2017;213(3):1022–9. https://doi.org/10.1111/nph.14265.

Publisher's Note