

Review

Population Genomic Approaches for Weed Science

Sara L. Martin ^{1,*} , Jean-Sebastien Parent ¹ , Martin Laforest ² , Eric Page ³ ,
Julia M. Kreiner ⁴ and Tracey James ¹

¹ Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada; jean-sebastien.parent@canada.ca (J.-S.P.); tracey.james@canada.ca (T.J.)

² Saint-Jean-sur-Richelieu Research and Development Centre, Agriculture and Agri-Food Canada, Saint-Jean-sur-Richelieu, QC J3B 3E6, Canada; Martin.Laforest@Canada.ca

³ Harrow Research and Development Centre, Agriculture and Agri-Food Canada, Harrow, ON N0R 1G0, Canada; Eric.Page@Canada.ca

⁴ Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada; julia.kreiner@mail.utoronto.ca

* Correspondence: Sara.Martin@Canada.ca; Tel.: +1-613-715-5406

Received: 16 August 2019; Accepted: 14 September 2019; Published: 19 September 2019



Abstract: Genomic approaches are opening avenues for understanding all aspects of biological life, especially as they begin to be applied to multiple individuals and populations. However, these approaches typically depend on the availability of a sequenced genome for the species of interest. While the number of genomes being sequenced is exploding, one group that has lagged behind are weeds. Although the power of genomic approaches for weed science has been recognized, what is needed to implement these approaches is unfamiliar to many weed scientists. In this review we attempt to address this problem by providing a primer on genome sequencing and provide examples of how genomics can help answer key questions in weed science such as: (1) Where do agricultural weeds come from; (2) what genes underlie herbicide resistance; and, more speculatively, (3) can we alter weed populations to make them easier to control? This review is intended as an introduction to orient weed scientists who are thinking about initiating genome sequencing projects to better understand weed populations, to highlight recent publications that illustrate the potential for these methods, and to provide direction to key tools and literature that will facilitate the development and execution of weed genomic projects.

Keywords: weeds; genomics; plant genome assembly; non-target site resistance; population genomics; genome scans; population genetics

1. Introduction

Biology is currently in the midst of a revolution caused by the advances in sequencing technology that allow us to examine genomes in detail [1]. Genomic information promises new insights for understanding the biology, evolutionary history, and adaptive potential in ways that were recently out of reach for laboratories studying organisms with genomes larger than model organisms (e.g., *Arabidopsis thaliana* (L.) Heyn. 135 Mb) [2–5]. Additionally, genomics at the population or species level are now possible in some species and will likely become practical for the majority of organisms in the near term. The huge potential of these advances has been exploited by some disciplines, such as those investigating bacteria [6,7], viruses [8,9] or humans [10–13], with greater alacrity than others. Notably, however, progress adopting genomic methods has been slow in weed science despite recognition of the power of these methods [14,15].

There are numerous impediments to a greater use of genomics in weed science. One of these elements is the lack of chromosome level reference genome sequences for weeds, as the majority

of sequencing efforts have been focused on crops. Genome sequences are foundational for many approaches and the relatively early availability of the human genome sequence [16], model organisms such as *Arabidopsis* [17] and many crop species [18] have been essential to the rapid progress in applying genomic approaches to a wide range of disciplines. This issue has been noted by the weed science community and efforts such as the International Weed Science Consortium have been initiated [14]. However, an additional impediment to using these rapidly developing and expanding set of techniques is a lack of familiarity among weed scientists. As a result, our aim here is to provide a brief primer and introductory “how to guide” and “why would you guide” relevant to weed science. We briefly review *de novo* genome assembly and annotation as these methods are often fundamental for further work. Then we focus on how genomic approaches can be used to answer three key questions: 1) Where do agricultural weeds come from and why are they weedy; 2) what genes underlie herbicide resistance (HR); and, more speculatively, 3) can we alter weed populations to become easier to control? We highlight what resources would be needed for success and provide illustrative examples from both weed science and the broader scientific literature.

2. Developing Weed Genome Sequences as a Fundamental Tool

While some genomic approaches do not require a draft genome for the species of interest, the majority of techniques do, or benefit from the availability of at least a rough draft. Sequencing plant genomes is easier than ever before with the decreasing cost of sequencing and the increasing ease with which tools such as genome assembly programs can be installed and used. However, genome assembly remains a challenge that will require a significant investment of time and resources for the majority of weed species [19]. Here we provide a brief outline of how to approach a *de novo* genome sequencing project and provide an initial introduction to the steps required and some tools that could be used as a starting point. We do not attempt to provide a comprehensive list of resources or tools and in every case, there are often numerous alternatives that may be better suited to a particular weed species or easier to install in a specific computing environment. Further, new tools are continuously emerging (and older ones submerging) in this quickly evolving area. Various databases of these tools have been compiled such as omictools.com and bioinformaticssoftwareandtools.co.in. Valuably, a recent review by Jung et al. [20] is comprehensive with recommendations on the computational resources needed to complete these assemblies.

2.1. What Is a Draft Genome?

A draft genome of a plant species is a haploid representation of a portion of the total DNA and genes. As such, it is a simplified and limited representation of the total information contained in the genome of the individual sequenced. It will lack information on allelic variation and portions of the genome, especially repetitive elements and material near the centromeres [21]. A draft genome is comprised of a group, often a large group (Table 1), of contigs that vary in size and represent the portions of the genome assembled from overlapping and joining the smaller pieces provided by the sequencing reads and is often presented in a multi-fasta file. These contigs can be assembled into larger fragments, scaffolds. Finally, scaffolds can be assembled, ordered, and oriented into pseudomolecules. At the larger end, pseudomolecules may represent chromosomes, chromosome arms or smaller features such as the chloroplast’s genome. In general, the fewer number of contigs an assembly has the better the assembly is considered. A metric used to compare continuity amongst genomes is the NG50 value. If one ordered all the contigs in an assembly from largest to smallest and added the length of each contig as you went down the list, the NG50 value would be the size of the contig when 50% of the species’ expected genome size was reached [22]. The N50 value is similar and more frequently reported, but as the assembly size is used instead of the expected genome size, it can’t be used to compare different assemblies even within species [22]. Drafts comprised of thousands of contigs can be sufficient for many purposes, including understanding evolutionary relationships among species, acting as a reference for studies of population biology, and for developing molecular identification

tools. To understand fine-scale patterns of selection, however, a chromosomal level assembly is more desirable, allowing for the most detailed analysis and inferences that draw on correlated shifts in allele frequencies. In cases where a closely related species has been assembled to the chromosome level and chromosome number is conserved, this may, by assuming synteny (preserved order), be used to position scaffolds into pseudomolecules representing a first guess of what the genome may look like. However, this level of information has rarely been achieved for non-model, non-crop organisms.

2.2. Preparing and Assessing Plant Material

Important initial steps to help ensure the success of a project are assessment of the plant material to understand the species' genome size and composition and carefully considering the starting material including finding lower ploidy individuals or reducing heterozygosity through inbreeding or other genetic manipulations such as creating a doubled haploid.

It is preferable to know the size of the genome before the start of a sequencing project. Several databases have compiled information on the genome size and chromosome counts for plant species (see Rice et al. 2015 for a list of resources). A particularly useful resource for genome size information is the Plant DNA C-value Database (cvalues.science.kew.org) hosted by Kew Royal Botanic Gardens [23]. Similarly, chromosome counts are available from the Index to Plant Chromosome Numbers www.tropicos.org/project/ipcn hosted by the Missouri Botanical Garden and the Chromosome Counts Database ccdb.tau.ac.il [24].

In the absence of information from these sources, or in cases where multiple chromosome counts or DNA contents have been reported, analysis by flow cytometry can determine the DNA content of the material of interest [25–28]. This is relatively inexpensive and straight forward if you have access to a flow cytometer and can take as little as a week for an experienced laboratory. Fresh tissue is co-chopped in a buffer with the tissue of a species with known DNA content (internal standard), nuclei are stained with a fluorophore such as propidium iodide, and peaks in fluorescence are produced as a result of excitation by the flow cytometer's laser. Then the position of the sample's peak and the known standard are determined by analysis of the resulting histogram with appropriate software (e.g., [29]). The DNA content of the samples is then determined using these relative positions and the DNA content of the standard. Generally, at least three individuals should be tested and each analyzed with three technical replicates across three days. This provides the full 2C DNA content of the plant's nuclei in picograms. The 1C DNA content can then be calculated by dividing this value in half and convert to Mbp by multiplying by 978 Mbp/pg [30]. Difficulties with DNA content determination with flow cytometry typically centre around finding an extraction buffer that allows for the production of narrow peaks and low debris levels (coefficient of variation < 5), including enough nuclei in the sample peak (>1000), finding an appropriate standard, and understanding the data when it is complicated by extra peaks from contamination or endopolyploidy [31,32]. Methods using an external standard should not be used as they are less accurate. This information can be compared to DNA content and chromosome counts for species in the same genus to make educated guesses about the chromosome count for the material of interest, but a conclusive determination of chromosome number requires either the counting of chromosome spreads or the use of more advanced chromosome sorting techniques [33].

Table 1. Metrics of continuity and completion for weed genome assemblies available from GenBank. This list was compiled by search GenBank [34] in May 2019 for species included on one of the following five lists of weeds: 1) Species with herbicide resistance maintained at weedscience.org by Heap [35], 2) the United States Department of Agriculture’s Federal Noxious Weed List [36], 3) Weeds of Nation Significance in Australia [37], 4) Weber and Gut’s list of weeds spreading in Europe [38], or 5) the Canadian Weed Seed Order [39]. Year is the year the assembly was submitted to GenBank, and as with assembly level, coverage, sequencing technology used and assembly method were recorded from the assembly information page on GenBank. The number of contigs (greater than 500 bp long), assembled genome size, N50 and NG50 were determined by QUAST (v. 5.0.2). The number of BUSCOs that were (C)omplete, complete and (S)ingle-copy, complete and (D)uplicated, (F)ragmented, or (M)issing were determined using the eudicotyledons_odb10 set of 2121 conserved genes and BUSCO version 3.0.2. Where we could not locate a published for the genome, we have reported the lead author as listed as having submitted the genome to GenBank. Note that additional weed genomes may be available on CoGe, Phytozome, and the European nucleotide database.

Common Name	Latin Name	Year	Level of Assembly	No. of Contigs	Est. Genome Size (Mbp)	Assembled Size (Mbp)	N50	NG50	BUSCOS (Percentage of 2121 Genes)					Coverage	Sequencing Technology	Assembly Method	Reference or Lead Submitting Author
									C	S	D	F	M				
Milkweed	<i>Asclepias syriaca</i>	2017	Scaffold	221,885	411 ¹	237	2555	NA ²	76	75	1	13	11	80.4	Illumina	Platanus SCUBAT	[40]
Winter Cress	<i>Barbarea vulgaris</i>	2016	Scaffold	7810	270	167	56,351	19,454	95	93	2	3	3	66.5	Illumina	Celera	[41]
Japanese Barberry	<i>Berberis thunbergii</i>	2018	Contig	11,815	1515 ¹	2241	397,058	654,137	88	30	57	3	9	104.8	PacBio	FALCON-Unzip	R. Bartaula
False Brome	<i>Brachypodium distachyon</i>	2018	Chromosome	11	355	271	59,130,575	59,130,575	80	76	4	6	15	9.4	ABI 3739	ARACHNE	[42]
Bird Rape	<i>Brassica rapa</i>	2017	Scaffold	70,673	485	386	3,737,062	2,395,810	98	80	18	1	1	212	Illumina PacBio	SOAPdenovo	[43]
Hemp	<i>Cannabis sativa</i>	2018	Chromosome	6653 ³	820	892	60,968,100	62,039,859	88	72	16	4	7	79	PacBio	FALCON	[44]
Shepherd’s Purse	<i>Capsella bursa-pastoris</i>	2017	Scaffold	8186	391 ¹	268	627,605	320,701	96	13	83	2	3	40	Illumina	Newbler Platanus	[45]
Horsetail Sheoak	<i>Casuarina equisetifolia</i> subsp. <i>incana</i>	2018	Scaffold	2936	340 ¹	301	1,020,118	894,734	97	93	4	1	2	546.9	Illumina PacBio	SOAPdenovo2 FALCON DISCOVER	[46]
Swamp Oak	<i>Cuarina glauca</i>	2018	Scaffold	39,787	340	283	964,272	627,004	97	93	5	1	2	890	Illumina	SOAPdenovo	[47]
Mandarin Orange	<i>Citrus reticulata</i>	2018	Scaffold	67,725	460	344	1,376,405	577,147	98	96	2	1	1	200	Illumina	Platanus	[48]
Horseweed	<i>Conyza canadensis</i>	2014	Contig	20,075	335	326	20,748	20,226	66	44	22	10	24	350	Roche 454 Illumina PacBio	Newbler SOAPdenovo CLC NGS Cell	[49]
Jute Mallow	<i>Corchorus olitorius</i>	2017	Contig	52,373	450	377	16,573	13,050	93	90	3	3	4	47.7	Illumina	Newbler	[50]
Muskmelon	<i>Cucumis melo</i>	2012	Scaffold	10,823	450	375	4,428,067	3,741,400	94	92	2	2	4	13.5	Roche 454 Illumina	Newbler	[51]
Globe artichoke	<i>Cynara cardunculus</i>	2018	Chromosome	8283 ³	1084	725	25,947,084	173,700	96	90	6	2	2	80	Illumina	AllPaths	[52]
Orchardgrass	<i>Dactylis glomerata</i>	2018	Scaffold	1,072,009	3327 ⁴	840	3242	NA ²	76	72	4	8	16	50	Illumina	SOAPdenovo	J. Li

Table 1. Cont.

Common Name	Latin Name	Year	Level of Assembly	No. of Contigs	Est. Genome Size (Mbp)	Assembled Size (Mbp)	N50	NG50	BUSCOS (Percentage of 2121 Genes)					Coverage	Sequencing Technology	Assembly Method	Reference or Lead Submitting Author
									C	S	D	F	M				
Carrot	<i>Daucus carota</i> subsp. <i>sativus</i>	2016	Chromosome	4826 ³	473	422	36,610,139	36,610,139	94	88	6	2	5	186	Roche 454 Illumina Sanger	SOAPdenovo GapCloser	[53]
Guinea yam	<i>Dioscorea rotundata</i>	2017	Chromosome	21	694 ¹	457	25,272,979	NA ²	83	78	5	3	14	100	Illumina	Allpaths-LG SSPACE Premium	S. Natsume
Barnyardgrass	<i>Echinochloa crus-galli</i>	2017	Scaffold	4113	1400	486	705,200	NA ²	89	26	63	2	10	170	Illumina PacBio	SOAPdenovo2 CANU	[54]
Paterson's curse	<i>Echium plantagineum</i>	2019	Chromosome	1091 ³	333 ¹	349	1,429,328	1,517,519	96	46	50	1	3	115	Illumina PacBio	MECAT LACHESIS	C.-Y. Tang
Common sunflower	<i>Helianthus annuus</i>	2017	Chromosome	1528 ³	3600	3028	178,899,001	174,509,413	89	80	9	3	8	100	PacBio	PBcR	[55]
Littlebell	<i>Ipomoea triloba</i>	2018	Chromosome	16	496 ¹	462	29,809,665	28,894,297	97	89	7	1	2	290	Illumina PacBio	SOAPdenovo2 SSPACE PBJelly, Pilon	[56]
Perennial Ryegrass	<i>Lolium perenne</i>	2016	Scaffold	666,180	2621 ¹	481	1361	NA ²	31	29	2	22	47	5	Illumina	CLC Genomic Workbench	[57]
Horsemint	<i>Mentha longifolia</i>	2016	Scaffold	190,876	400	353	3915	3044	58	52	5	20	22	33	Illumina PacBio	MaSuRCA	[58]
Amur silver grass	<i>Miscanthus sacchariflorus</i>	2018	Chromosome	105,321 ³	2513 ¹	2075	37,709	24,189	49	41	8	17	33	60	Illumina	ABYSS SOAPdenovo2	J. De Vega
Longstamen Rice	<i>Oryza longistaminata</i>	2014	Scaffold	9688	782 ¹	362	30,401,905	NA ²	86	80	6	4	10	52.5	Illumina	SOAPdenovo2	C. Brian
Red Rice	<i>Oryza punctata</i>	2014	Chromosome	12	586 ¹	394	31,244,610	28,494,620	81	74	7	6	13	130	Roche 454 Illumina	AllPaths	R. A. Wing
Brownbeard Rice	<i>Oryza rufipogon</i>	2015	Scaffold	3818	450 ¹	339	27,785,585	26,200,591	83	76	6	5	13	120			Q. Zhao
Rice	<i>Oryza sativa</i>	2019	Chromosome	367 ³	489 ¹	415	28,085,715	26,003,091	88	81	6	3	9	148	PacBio	CANU	L. Wang
Broomcorn Millet	<i>Panicum miliaceum</i>	2018	Chromosome	466	923	848	48,259,421	45,112,342	83	25	58	4	13	160	Illumina PacBio	CANU	[59]
Opium Poppy	<i>Papaver somniferum</i>	2018	Chromosome	34,381 ³	2870	2716	204,470,928	180,516,484	95	29	65	1	4	239	Illumina PacBio ONT	DeNovoMAGIC FALCON	[60]
White Poplar	<i>Populus alba</i>	2019	Contig	6087	508 ¹	707	248,703	390,844	95	52	43	1	3	130	Illumina PacBio	SMARTdenovo	[61]
Algarrobo blanco	<i>Prosopis alba</i>	2019	Contig	4454	391 ¹	500	237,044	357,710	70	49	21	3	27	30	PacBio	CANU	W. Kong
Wild Radish	<i>Raphanus raphistrum</i>	2014	Contig	64,732	515	254	10,333	NA ²	95	82	12	3	2	47	Roche 454 Illumina	ABYSS, Newbler, Celera Assembler, Minimus2	[62]
Radish	<i>Raphanus sativa</i>	2017	Chromosome	44,239 ³	573	383	35,166,889	26,198,371	96	82	14	2	1	225	Illumina	SOAPdenovo2	[63]
Japanese Rose	<i>Rosa multiflora</i>	2017	Scaffold	83,189	711	740	90,830	95,085	91	66	25	4	5	327	Illumina	SOAPdenovo2 GapCloser	[64]
Wild Sugarcane	<i>Saccharum spontaneum</i>	2018	Chromosome	15,303 ³	1565 ¹	3133	91,359,291	109,189,819	78	20	58	5	17	90	Illumina PacBio	CANU HiC	[65]

Table 1. Cont.

Common Name	Latin Name	Year	Level of Assembly	No. of Contigs	Est. Genome Size (Mbp)	Assembled Size (Mbp)	N50	NG50	BUSCOS (Percentage of 2121 Genes)					Coverage	Sequencing Technology	Assembly Method	Reference or Lead Submitting Author
									C	S	D	F	M				
Rye	<i>Secale cereale</i>	2017	Scaffold	1,581,707	7900	1685	2200	NA ²	66	62	4	13	21	50	Illumina	CLC Assembly Cell, CarmA	[66]
Green Foxtail	<i>Setaria viridis</i>	2019	Chromosome	14	782 ¹	396	46,702,114	35,460,007	81	75	6	6	13	118	PacBio	MECAT	P. Huang
White Champion	<i>Silene latifolia</i>	2018	Scaffold	319,506	2640 ¹	1185	11,019	NA ²	68	66	4	13	18	40	PacBio	SOAPdenovo2, CLC, PBJelly, SSPACE	[67]
Milk Thistle	<i>Silybum marianum</i>	2016	Contig	258,575	792 ¹	1478	6967	NA ²	38	33	6	8	54	96	Illumina PacBio	Celera Assembler	Y. Lv
Sorghum	<i>Sorghum bicolor</i>	2017	Chromosome	869 ³	730	709	68,658,214	68,658,214	86	80	5	4	10	8	Illumina Sanger	ARACHNE	[68]
Stinkweed	<i>Thlaspi arvense</i>	2015	Scaffold	6768	539	343	140,815	NA ²	98	97	2	1	1	80	Illumina PacBio	CLC NGS Cell	[69]

¹ When not reported by the authors, we have estimated based the genome size based on the genome size available from Kew's C-DNA value database (see Section 2.2). In some cases, this has resulted in an estimate smaller than the assembled genome size. ² In cases where the genome assemble size is not sufficiently higher than half of the expected genome size, an NG50 cannot be calculated (see Section 2.1). ³ In some cases chromosome-level genome assemblies have pieces left over and these increase the number of contigs included in the assembly files beyond the expected chromosome number. ⁴ Genome size estimate from DNA content analysis in Creber et al. [70].

Producing chromosome spreads is generally more accessible than chromosome sorting, but requires a significant amount of time, especially if the species' chromosomes are small or numerous. A pair of highly helpful videos on the technique, produced by the Beck Laboratory are available as an introduction (www.youtube.com/watch?v=iXqni6knH5A&t and www.youtube.com/watch?v=xVV4qBfSQLs&t) [71,72]. Several methods that inhibit spindle formation and increase the accumulation of metaphase cells can be used to facilitate chromosome counts. These include pre-treating material with pressurized nitrous oxide (NO₂), incubation in ice cold water, or exposing the cells to chemical inhibitors such as 8-hydroxyquinoline or colchicine [73]. For example, for a mitotic preparation, NO₂ pressurized to 8–10 atm (160 psi) can be applied for several hours to 1 cm long root tips in water using a specially constructed air sealed, iron pressure chamber with the regulator and hoses of the correct composition needed to attach and deliver NO₂ [74]. The water is then removed and replaced with fresh Carnoy's fixative for storage at 4 °C. Samples are then washed twice with distilled water and 1x citric buffer respectively. This buffer is replaced with enough 0.3% pectolytic enzyme solution [75] to ensure the material is fully submerged and incubated at 37 °C for 60 min. Digested root tips should form into a cell suspension when they are tapped with dissecting needles on a slide. If clumps form, or cells do not separate, incubation in the enzyme solution should be increased. Once cell suspension has been created, a drop of orcein stain [73] can be added [73], the drop carefully spread, and a coverslip placed on top. Then the slide is heated and squashed between filter paper using thumb pressure, ensuring no slippage. The slide can then be examined with a phase-contrast microscope for the quality of chromosome spread and count. If cytoplasm covers the nuclei then pepsin treatment may be effective [75]. Obtaining a good spread that will allow for certainty in chromosome number will generally take patience and practice.

An alternative to flow cytometry for determining genome size is to complete a k-mer plot of Illumina short read data (Illumina, San Diego, California, USA) [3] using a tool such as KmerGenie (kmergenie.bx.psu.edu) [76] or Jellyfish (www.cbcb.umd.edu/software/jellyfish), however it is preferable to have an estimate independent of the reference read data itself [77] (Figure 1). Following data generation with Jellyfish a script can be written in R to visualize the data or the data can be easily visualized using the website GenomeScope (qb.cshl.edu/genomescope). This data can also provide an indication of heterozygosity and can be used to determine the amount of the genome comprised of repetitive elements using tools such as RepeatExplorer (repeatexplorer.org) [78,79].

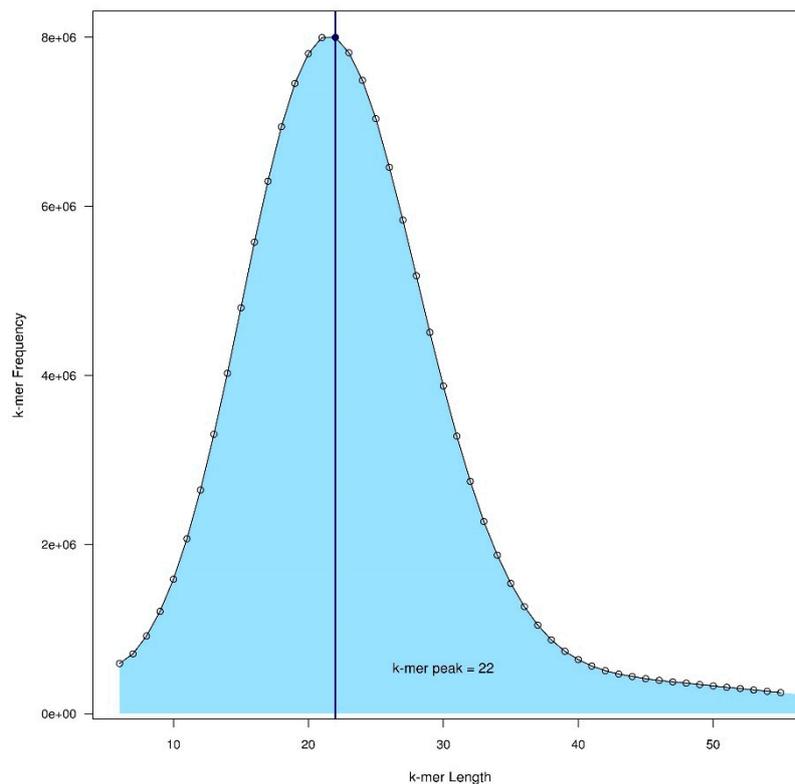


Figure 1. Plot of k-mer frequency by length produced for *Camelina neglecta* J.Brock, Mandáková, Lysak & Al-Shehbaz produced using Jellyfish and visualized using R. The position of the peak at a k-mer length of 22 is used to calculate genome size based on the area under the curve as represented by the light blue region. Here the genome size estimated is 248 Mb, while flow cytometry estimates indicate a genome size of 264 (± 9) Mbp [80].

While the addition of long-read technology is making the assembly of highly heterozygous and repeat-rich genomes more feasible, genome assembly can be simplified by reducing heterozygosity and repetitive elements. In species that are self-compatible, repeated self-pollination can do both and result in a less redundant and more contiguous assembly [81]. In outcrossing or dioecious species or species with strong inbreeding depression reducing variation can be more difficult, requiring strategies such as repeated full sibling mating. Doubled haploids, generally produced via tissue culture of either male or female gametophytes, can solve this problem by completely eliminating heterozygosity, but are also a significant challenge and investment of time [82–84].

For genome sequencing, the 1C DNA content is perhaps the most important piece of information for designing the sequencing strategy, determining the quantity of sequencing required, and providing hints as to the species' degree of polyploidization or genome size inflation resulting from repetitive element proliferation.

Additional challenges await groups that wish to assemble genomes which have undergone recent or ancient polyploidization, which are notoriously more difficult to assemble, though long reads are making these genomes increasingly tractable. Many successfully sequenced crops fall in this category and specific strategies have been developed to assemble these genomes (reviewed by [85]). However, when a weed species is variable for ploidy the most feasible approach would be to select an individual with the lowest ploidy available for sequencing. However, the conclusions about the species' population genetics that are drawn from this genome would only be applicable to populations with this cytotype. In any case, a vouchered record of the material used for DNA extraction should be created and submitted to an herbarium to provide documentation of the species that has been sequenced [86].

2.3. DNA Extraction

Extraction of DNA of sufficient quality and quantity can be a surprisingly difficult hurdle. Technologies such as Pacific Biosciences' (PacBio) single molecule real-time (SMRT) sequencing (Pacific Biosciences, Menlo Park, California, USA) and Oxford Nanopore Technologies' sequencing systems (Oxford Nanopore Technologies, Oxford, UK) require high molecular weight (HMW) DNA at a high concentration (e.g., 10 µg with an average size of 30–50 kbp for PacBio) [87]. This genomic HMW DNA needs to have little evidence of shearing, be free of contamination from protein, RNA, or polysaccharides and a 260/280 nm absorbance ratio of approximately 1.8–2.0. This is not always simple to achieve and time may need to be devoted to optimizing the DNA extraction protocol.

We have observed that the method of grinding the plant tissue appears to be the most critical step in obtaining HMW DNA with little shearing (Martin, unpublished). While many protocols suggest using bead mills with either ceramic, metal beads and/or sand, using the least time and speed reduces shearing. We have found that grinding tissue in 2 mL tubes with plastic pestles on dry ice, using wide bore tips, minimizing vortexing and pipetting will limit shearing and help ensure recovery of HMW DNA. Commercial kits are convenient and remove contaminants, but often an insufficient amount of DNA is obtained from a single extraction. However, multiple extractions can be pooled and concentrated to obtain the HMW at a sufficient concentration.

When sufficient tissue is available, many genome sequencing projects (e.g., [44,54,88,89]) have found success with variations on the traditional hexadecyltrimethylammonium bromide (CTAB) based method, described by Doyle and Doyle [90]. These methods often use a large quantity (g) of plant tissue ground in liquid nitrogen with a mortar and pestle. Many modifications of this original protocol are available, including Healy et al.'s [91] protocol for plants with large amounts of phenolics and polysaccharides. These compounds can inhibit downstream library preparations and are particularly important to eliminate. If required, further purification can be done with additional ethanol precipitations or magnetic beads (Agilent, Santa Clara, California, USA). For example, a strategy to prepare fragments for sequencing is to shear the DNA into large fragments of 20 kb in size using g-TUBES (Covaris, Woburn, MA, USA) and then selecting fragments of appropriate size with an apparatus such as the Blue Pippin (Sage Science, Beverly, MA, USA). In addition, some laboratories have found specially designed tips, such as Qiagen Genomic Tips (Qiagen, Hilden, Germany) to be helpful during preparation of the samples. Other technologies such as the Short Read Eliminator Kit (Circulomics, Baltimore, MD, USA) can be used to optimize sequencing by removing shorter fragments. Following extraction, DNA integrity and concentration need to be assessed. A variety of tools exist to complete these steps including the TapeStation or Bioanalyzer system (Agilent Genomics) [87]. However, it has been noted that DNA quantities should be measured on a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) or similar as Nanodrop (Thermo Fisher Scientific) can overestimate quantity [87].

2.4. Sequencing Strategies

Assembling a genome using large pieces is much easier than using small pieces. Therefore, the majority of sequencing projects now combine long read (e.g., PacBio or ONT) and short read data. Long reads, which generally average 10 kb or more in length, make assembling plant genomes comparatively easier and general result in a more contiguous assembly. Genome assembly is sensitive to repeated sequences and these can only be resolved if the sequencing technology spans the regions. However, the error rate for long reads maybe as high as 15% and therefore require greater depth (30× per haploid genome, see below) to allow a consensus to be called from the data [20]. While short read Illumina data is unable to resolve long repeats, it has higher accuracy and can be used to correct long read data [4] either before or after assembly to improve the accuracy or completeness of the genome [20].

The recommended coverage for genome assembly varies from 40× to 60× at a minimum. For example, Li and Harkness [3] suggest 40–50× and Del Angel et al. [4] and Jung et al. [20] suggest

a minimum of 60× for small, inbred, diploid genomes. Coverage is generally estimated based on the Lander-Waterman equation [92] as read length multiplied by read number divided by the haploid genome size for the species. Perhaps more simply for project planning, the amount of sequencing data needed for a project can be calculated by multiplying the estimated size of the plant's haploid genome by the coverage needed. However, it is important to note that coverage will be reduced by quality control and filtering steps compared to the raw coverage. Additionally, the coverage will not be uniform across the nuclear genome. For example, up to 20% of the raw data may be DNA from the chloroplast resulting in relatively deep coverage of the relatively small chloroplast genome, but less coverage of the nuclear genome [93]. After generating sequence data, there are generally five, often iterative, steps before the “final” genome is ready for downstream analysis: 1) Data assessment and filtering, 2) assembly (often by multiple assemblers), 3) error correction and polishing, 4) scaffolding and/or the placement of scaffolds on chromosome sized pseudomolecules, and 5) annotation.

2.5. Data Assessment, Correction and Filtering

Before starting with the assembly process, it is advisable to assess the quality of the sequencing data and filter the reads based on this quality. However, some assemblers integrate quality filtering and correction as early steps in their assembly process and additional steps with alternative software may or may not improve the final assembly. Read length can also be a consideration as, for example, some long read assemblers will refuse to work if reads shorter than 500 bp are included in the input data. The software FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) provides a summary of quality parameters that is very helpful to assess the quality of short or long read data: Average per base quality, per tile quality, per sequence quality, per base content, per sequence GC content, per base N content, sequence length distribution, sequence duplication level, overrepresented sequences, adapter content and k-mer content. Overall quality of long read data can be also assessed with tools such as Nanoplot (github.com/wdecoster/NanoPlot) [94]. Correction of long read data with short reads can be done prior to assembly with tools such as LoRDEC (www.atgc-montpellier.fr/lordec) [95]. Filtering can be done with a variety of tools available online such as Trimmomatic (www.usadellab.org/cms/?page=trimmomatic) [96]. This type of software will generally remove reads or regions in the reads that are below a certain quality threshold as well as sequencing adapters or the “bar codes” of specific sequences that allow for identification of particular reads following multiplexing. Many custom scripts for filtering raw data can be found online (e.g., `filter_fastq.py` github.com/nanoporetech/fastq-filter/blob/master/filter_fastq.py). Users will want to apply the principle of *caveat emptor* when using these scripts, but they can provide invaluable tools.

2.6. Assembly and Assessment

Genome assemblers typically use either short or long read data as input. Short read assemblers have a longer history and many are designed with smaller bacterial or viral genomes in mind. However, because of their longer history, several of the programs that can handle larger genomes have also had extensive work to reduce the amount of computational resources they need such as ABySS 2.0 (www.bcgsc.ca/platform/bioinfo/software/abyss/releases/2.0.0) [97] and SOAPdenovo2 (github.com/aquaskyline/SOAPdenovo2) [98]. In our experience, two genome assemblers that use long read data that are relatively easy to install and use with strong documentation and community support are CANU (canu.readthedocs.io/en/latest) [99] and FALCON (pb-falcon.readthedocs.io/en/latest) [100]. CANU, in particular, appears to be a common choice (Table 1), perhaps because of the clarity of its documentation and recommendations on which parameters (e.g., `correctedErrorRate` and `minOverlapLength`) are the most likely to improve the outcome of the assembly. This type of guidance is very helpful as the key parameters for tuning software to a particular species are not always apparent, resulting in an overwhelming number of parameters that could be adjusted. However, when in doubt and lacking documentation, this information can also be gleaned from other users' experience documented in discussion groups for the particular tool. Hybrid assemblers, that use

both short and long read data, such as SPAdes (github.com/ablab/spades) [101], and Platanus-allee (platanus.bio.titech.ac.jp/platanus2 the recent replacement of Plantanus) [102] are available and assembly strategies that merge the results of multiple assemblers have also been used (e.g., [89]).

Once an assembler has completed a draft assembly of the genome, the challenge is determining how “good” the assembly is [19]. The definition of good can depend on the eventual use of the genome and includes parameters such as how contiguous (how many pieces is the genome in) the assembly is, how much of the genome was assembled and whether the assembly contains the expected genes. Often the first tool applied following genome assembly is QUAST (quast.sourceforge.net/quast), which provides a quick summary of the genome including the number of contigs, the total length of the genome as assembled, the N50, and, if the expected genome size is included the NG50 values. This gives an indication of contiguousness and the size of the assembly. BUSCO (busco.ezlab.org) [103,104] is frequently used as a quantitative measure of the completeness of a genome as it indicates whether the shared single copy genes expected in the genome are present—that is how much of the gene space has been captured and assembled. BUSCO indicates how many and which of these are complete and single copy, complete and duplicated, missing or fragmented (Table 1). Finally, BlobTools (blobtools.readme.io/docs) [101] can be used to determine if the assembled sequences are DNA from the expected organism or from contaminating organisms through taxonomic partitioning of the genome. This tool requires the draft genome sequence, a hit file created by BLASTn (blast.ncbi.nlm.nih.gov/Blast.cgi) [105] using the MegaBLAST option [106], a depth file created with a tool such as BWA-MEM (bio-bwa.sourceforge.net) [107], and the raw data used to assemble the genome sequence. After processing this information BlobTools creates a visual indication of which organisms are most closely related to the draft genome (Figure 2). If there is substantial contamination, this information to further filter the raw data for reassembly without the contaminating sequences.

2.7. Polishing

Polishing a genome can lead to significant improvements in the completeness of the genome as assessed by BUSCO and some tools will use short read data to call a consensus SNP, correct indels (insertions and deletions that are common in log read data) and misassembled contigs. Pilon (github.com/broadinstitute/pilon) [108] uses the assembled genome and one or more files containing the alignment of sequencing reads such as mate pairs, paired ends or unpaired sequences to the draft assembly. The program’s output includes the files needed for visualizing the changes to the genome using tools such as the Integrative Genomics Viewer (IGV software.broadinstitute.org/software/igv/) [109] and can generate information on the variation with genome sequence. PacBio has developed the tool GenomicConsensus (github.com/PacificBiosciences/GenomicConsensus), which uses mapped PacBio reads to generate a consensus, while Nanopolish (nanopolish.readthedocs.io/en/latest/index.html) has been developed for use with ONT data. In comparison, RACON (github.com/isovic/racon) can be used with either short read or long read data [110].

2.8. Scaffolding

Traditionally, the ordering and orientation of contigs into scaffolds has often relied on the labor intensive and expensive use of fluorescent in situ hybridization of bacterial artificial chromosomes (BACs) and segregating F2 populations that allow for mapping the position of the sequences. More recent methods: Chromosome conformation capture techniques (Hi-C), optical mapping techniques (Bionano) and 10x Genomics Chromium™ Systems can produce data that can be generated and applied to verify the assembly and generate scaffolds with less time and effort [3]. Chromosome conformation capture (3-C) has been a commonly used technique in molecular biology to map chromosomal interactions. It uses a process where genomic DNA is first digested and then ligated in conditions that preserve the 3D organization of the genome to allow the joining of distant sequences that find themselves to be in proximity. Using deep sequencing, the high throughput version of the technique (Hi-C) produces a genome-wide map of proximity contacts between all the different loci. Since the

frequency of occurrence of such contacts is based on proximity, with intrachromosome contacts most common and the probability of contacts decreasing with distance, the technique can readily be used for scaffolding contigs [111]. If the analysis of this proximity data is not completed by the provider using proprietary software, once the paired end data has been mapped to assembled contigs, software such as SALSA (github.com/machinegun/SALSA) [112] can use the information to break misassembled contigs and scaffold the genome. FALCON-Phase (github.com/PacificBiosciences/pb-assembly) has also integrated the use of Hi-C data into the FALCON assembly pipeline through a collaboration between PacBio and Phase Genomics (www.phasegenomics.com) [113]. Phase Genomics is a USA based company that can provide kits for HI-C library preparation and bioinformatics support in the use of this data scaffolding of a *de novo* genome with their proprietary software Proximo. Additionally, they provide helpful advice on how to work with Hi-C data generated by their protocols (phasegenomics.github.io/2019/09/19/hic-alignment-and-qc.html). Recently, chromosome level assemblies of black raspberry (*Rubus occidentalis* L.) [114], an ornamental amaranth used by ancient civilizations in South and Central America as a grain crop (*Amaranthus hypochondriacus* L.) [115], and broomcorn millet (*Panicum miliaceum* L.) [59], genomes have been completed using Hi-C data and PacBio data.

Bionano Genomics (San Diego, CA, USA, bionanogenomics.com) contributes to scaffolding by optically mapping specific sequences distributed across the genome. Briefly, high molecular weight DNA is extracted, up to chromosome arm lengths, and labeled at specific sequence motifs for imaging and identification. The DNA molecule is then linearized onto a flowcell where a gradient of micro- and nano-structures gently unwinds and guides DNA into NanoChannels where it is imaged by a high resolution camera. The DNA fragments with similar motif-specific label patterns are assembled together to recreate a whole genome map assembly. This data can be used in a hybrid assembly to scaffold contigs obtained through sequencing of the genome. It can be used to identify regions that are incorrectly assembled or where structural variants can be found. This approach was recently used in the improvement of wheat's hexaploid genome assembly [116] and the large *Sorghum* genome [117].

An alternative approach is used by 10x Genomics Chromium™ System (www.10xgenomics.com). DNA molecules are divided into small sets and provided with an identifying barcode before being sequenced. This provides linked reads that are unlikely to represent the same region from homologous chromosomes. This technique is particularly useful in genomes that are highly heterozygous and/or polyploid because it allows the genome information to be phased, that is the two haplotypes can be distinguished, and it can prevent the collapse of sequence from homologous chromosomes in polyploids. This technique was recently used in the sequencing of the octaploid strawberry genome (*Fragaria X ananassa*) [118].

An additional option when a related species with a chromosome-level genome sequence is available, is that this information can be used to create reference based assembly with chromosome-level resolution. However, this method would bias the assembly to more closely resemble that of the relative and will, for example, lack chromosome scale rearrangements. One option for pursuing this route, MeDuSa [119] (github.com/combogenomics/medusa/releases), can use one or more closely related genomes for generating a chromosome-level draft.

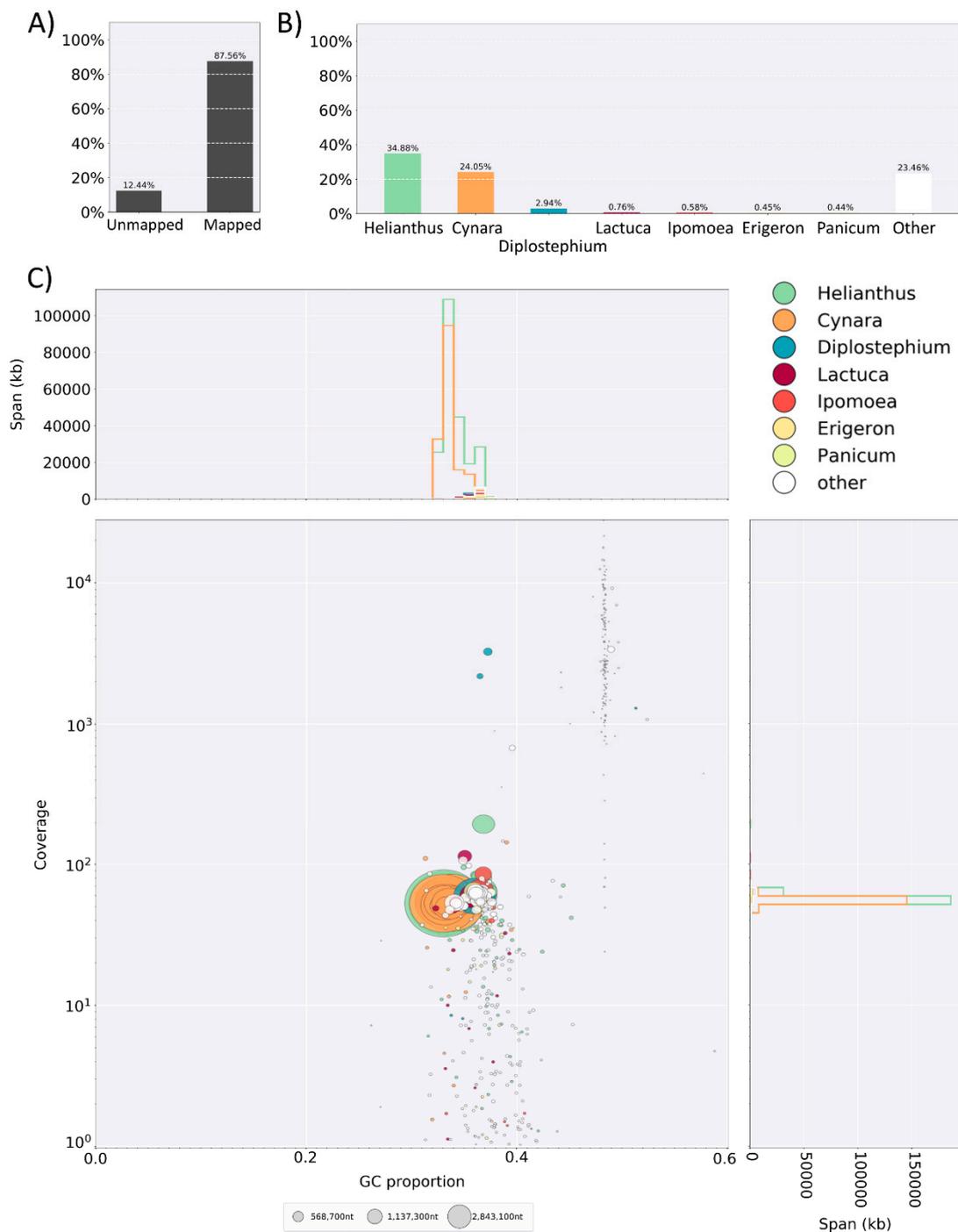


Figure 2. Blobplot generated for *Conzya canadensis* (Asteraceae) draft genome assembly showing the genera with the closest similarity to the sequenced genome (Laforest, Martin, and Page unpublished data). The first panel (A) indicates the percentage of reads that were mapped and the second panel (B) shows the taxonomic break down of hits at the taxonomic level requested. In this case the majority of hits are from other genera from the Asteraceae. The program generates a text file with more detailed information. The three part third panel (C) shows histograms for the proportion of G and C bases in the sequence which typically varies among species (top) and coverage (right) weighted by the cumulative length of sequences in each bin. The main panel has circles colored by taxonomic affiliation positioned on the x-axis by the GC proportion and on the y-axis by coverage within the raw data which gives a sense of the relative concentration of the sequences in the DNA sample.

2.9. Gene Prediction and Annotation

Once a genome sequence of adequate quality has been produced, genes and other genetic elements such as transposons need to be identified. Gene prediction software such as AUGUSTUS (bioinf.uni-greifswald.de/augustus) [120,121] can be used to locate potential coding sequences along the genome sequence. This software has been improved over the years, starting from entirely ab initio gene prediction to include evidence-based discovery using expressed sequence tag (EST) sequences, RNASeq data (by way of hints) and with protein multiple sequence alignments. Repeated elements such as transposable elements (retrotransposons and DNA transposons), tandem or inverted repeats, can be located in the genome with software such as RepeatMasker (www.repeatmasker.org), RepeatFinder (www.cbcb.umd.edu/software/RepeatFinder) [122], or the recently developed Generic Repeat Finder (GRF) [123]. Additionally, there are a host of software packages and resources designed to detect and annotate specific types of transposable elements including SINE_scan (github.com/maohlzj/SINE_Scan) [124] for detected short interspersed nuclear elements (SINEs), the P-Mite database (pmite.hzau.edu.cn) [125] for finding miniature inverted-repeat transposable elements, and HelitronScanner (sourceforge.net/projects/helitronscanner) [126] for detecting helitrons—rolling circles that often capture gene sequences leading to gene duplication.

It is useful to know what the product of identified gene sequences code for and tools have been designed to assign gene ontology—information on a gene's product's molecular function, location and role (GO, geneontology.org) using standardized language. One of the most ubiquitous tools used is the basic local alignment search tool (BLAST) [105] in conjunction with the Genbank [34] databases to assign putative functions through shared identity or similarity of the translated gene product. Blast2Go (www.blast2go.com) [127] is a tool with a subscription fee that can automate this process. Free software packages are also available including the widely used Maker-P (www.yandell-lab.org/software/maker-p.html) [128] as pipeline designed to make the annotation of plant genomes more accessible to new groups and incorporates many of the software packages mentioned above and has extensive documentation and tutorials.

2.10. Examples: Three Recently Sequenced Weed Genomes

Given the wide variety of sequencing strategies and tools that can be employed (or not) at each stage of genome assembly it is unlikely that any two projects have followed the same path to a final assembly. Further, as noted by Del Angel et al. [4], it is important to set goals at the beginning of a project for how contiguous and complete the genome sequence needs to be for the specific project, otherwise the iterative process of analysis and reanalysis with alternative tools can be endless. Given the complexities of genomes (e.g., [129]) and how this complexity is reduced in a genome assembly, it may be helpful to consider a modification of George E. P. Box's aphorism that all genome sequences are wrong, but some are useful. As examples of how these techniques and programs have been applied to weeds, we briefly summarize the methods and outcomes of three recent sequencing projects of two diploids, kochia (*Kochia scoparia* (L.) Schrad. also called *Bassia scoparia* (L.) A.J.Scott), common waterhemp (*Amaranthus tuberculatus* (Moq.) Sauer), and a hexaploid species, barnyard grass (*Echinochloa crus-galli* (L.) Beauv.).

For kochia, a plant with a genome size of approximately [89] 1Gbp ($2n = 2x = 18$), DNA for sequencing was extracted from a glyphosate susceptible inbred line using a modified CTAB protocol. They sequenced three Illumina libraries, one paired end and two mate-pair libraries using three HiSeq lanes and used 12 PacBio SMRT cells. They then assembled and merged two assemblies into a final assembly for analysis. For the first assembly, they used the paired end data and the program Proovread (github.com/BioInf-Wuerzburg/proovread) [130] to correct the PacBio reads, which were then assembled with Canu. For the second assembly, ALLPATHS-LG (software.broadinstitute.org/allpaths-lg/blog) [131] was used to assemble all the Illumina data and scaffolding was completed using the PacBio reads and PBJelly (sourceforge.net/p/pb-jelly/wiki/Home) [132]. They then used the GARM Meta assembler (garm-meta-assem.sourceforge.net) [133] to merge the genomes. This final

711 Mbp assembly consisted of 19,671 scaffolds and had an N50 of 62 kb. Completeness as indicated by BUSCO, using the eudicotyledons odb10 dataset, was estimated at 70.3%. Kochia's sequence was then annotated using the WQ-Maker pipeline transcriptome data from kochia and expressed sequence tags for kochia's family, the Chenopodiaceae, from the National Center for Biotechnology Information (NCBI www.ncbi.nlm.nih.gov). Then then used BLASTN and BLASTP to predict genes and proteins and RepeatMasker to search for repetitive elements.

In the case of common waterhemp, a species with a genome size of approximately 676 Mbp ($2n = 2x = 32$), DNA from a single female plant was extracted using a modified CTAB protocol and sequenced with both PacBio reads, 15 SMRT cells, and one Illumina HiSeq lane of 150 bp paired end library reads [88]. The long read data provide $87\times$ coverage and was assembled using Canu and then polished with the short read data using Arrow and Pilon. This resulted in a final genome assembly size of 663 Mbp consisting of 2,514 contigs and an N50 of 1.7Mb. The assembly contained 88% of BUSCO's Embryophyta's genes. The program REVEAL (github.com/jasperlinthorst/REVEAL) [134] was then used to produce 16 pseudomolecules using the chromosomal level genome assembly of the cereal crop species *Amaranthus hypochondriacus* L. Both this finished genome and the assembly used to create it were annotated using the MAKER pipeline (yandell-lab.org/software/maker.html) following identification and masking of repetitive elements with RepeatModeler and RepeatMasker.

Barnyard grass has an estimated genome size at 1.4 Gbp based on flow cytometry data and K-mer analysis [54] and a chromosome count of $2n = 6x = 54$. DNA was extracted for sequencing from a plant collected from a rice paddy using a CTAB protocol. They sequenced the 48 SMRT cells of PacBio for long read data and both paired end and mate pair Illumina libraries using HiSeq runs. This level of sequencing effort resulted in $171\times$ coverage of the genome. The short read data was assembled with SOAPdenovo2, scaffolded with OPERA-LG (sourceforge.net/p/operasf/wiki/The%20OPERA%20wiki) [135], and then gaps in this assembly were closed with GapCloser from SOAPdenovo2. The long read data was assembled with Canu and used to fill gaps in the short read assembly with PBjelly. The draft genome produced was 1.27 Gbp in length with an N50 of 1.8 Mbp. The authors used BUSCO and determined that 95.5% of the core eukaryotic genes were complete. RepeatModeler and RepeatMasker were used to find and mask repetitive elements. Then they used transcriptome data and three programs to predict genes GeneMark.hmm (exon.gatech.edu/GeneMark) [136], Fgenesh (www.softberry.com) [137], and AUGUSTUS.

3. Current Application: What Are Agricultural Weeds and Where Do They Come From?

Harlan and deWet defined weediness as "an adaptive syndrome which permits a species or variety to thrive and become abundant and difficult to eradicate within areas of human disturbance" [138]. Under this definition, crops are the result of intentional selection for vigor and fertility in the agricultural environment and weeds are the unintentional result [139]. A classic example of this is crop mimicry, where weeds have been selected by agricultural practices such as hand weeding to closely resemble a crop species [140]. This includes species such as false flax (*Camelina sativa* (L.) Crantz), which looks like, has similar time to maturity, and similar seed size to varieties of cultivated flax [141,142], and rice-mimicking varieties of barnyard grass [140]. A more pressing example is the evolution of HR (see Section 4) [143]. This second example illustrates, that as a group, weeds represent multiple independent origins of weediness and numerous examples of rapid adaptive evolution that present an opportunity not only to co-opt these adaptations for crop improvement or guide changes in agricultural practices to slow or thwart this evolution [144], but to provide fundamental insights into evolution [145]. Agricultural weed populations can be selected from populations adapted to natural disturbance regimes or from populations selected for these characteristics as crops, from populations of wild crop relatives, or from hybrids between the two [141,146–148]. Similarly, specific traits that contribute to adaptation to the agricultural environment, including alleles conferring HR, are selected within those populations. These origins and the loci underlying adaptive traits can be elucidated by examining genomic variation with weed populations.

3.1. Detecting the Signatures of Demographic Change and Selection on the Genome

Demographic and selective events change the patterns of variation across the genome, leaving a record of these processes. In weed populations, demographic and selective events may be closely intertwined as artificial selection from weed control measures can drastically change population size and composition. For example, weed populations might undergo rapid declines in population size (bottlenecks) resulting from herbicide application followed by population expansions after the evolution of HR, or the introgression of HR genes from one population into another. These processes can be difficult to disentangle from each other, as well as from patterns related to the variable recombination rate across the genome. However, demographic processes generally leave a signature across the entirety of the genome, while selection leaves a signal localized to the genes that confer higher fitness under the given environmental regime.

Over time, adaptation of a population to its specific environment and associated demographic events lead to divergence in allelic composition across the genome relative to other populations. This divergence leads to population structure and can be used to infer the past history of the sample, with populations sharing more similar allele frequencies more likely to share a recent evolutionary history. When a species exhibits population structure, we can assign individuals to recent common “ancestral populations” that can provide clues to their origin. This is often the basis of human ancestry assignment through home DNA tests, where your genotyping results are compared to the frequency of alleles across the globe to determine which geographic region contains the highest proportion alleles similar to those comprising your genotype [149,150]. Population structure can also provide evidence of hybridization and introgression when individuals show the signal of a mixed affinity to populations or species (admixtures). Again, this is similar to the assignment of percentage affiliation to different groups in human ancestry tests.

Population structure can be estimated at many hierarchical levels, from individual, to subpopulation, and across longer timescales at the phylogenetic level (e.g., STRUCTURE (web.stanford.edu/group/pritchardlab/structure.html) [151], AMOVA [152], and TREEMIX (bitbucket.org/nygcresearch/treemix/wiki/Home) [153]). While these methods aim to cluster individuals into discretely structured groupings, allele frequencies may instead continuously vary across space [154]. This may be especially likely for a recently expanded species due to serial bottlenecks and expansions, or along clines in latitudinal or environmental gradients where there is limited opportunity for long distance dispersal [155,156]. However, methods have been developed to test whether a population is more likely to showing continuous or discrete population structure [157]. In these cases, a model free approach such as principal component analysis may help to clarify population structure [158]. These data can also be used to infer past demographic processes using modelling approaches that allow estimation of parameters including ancestral population size, the number and timing of bottlenecks, time since divergence between populations, ancestral and contemporary levels of gene flow, and contemporary effective population sizes. Demographic modelling has been widely implemented to infer the history of sampled populations including *δαδι* (bitbucket.org/gutenkunstlab/dadi/src/master/) [159] and FastSimCoal (cmpg.unibe.ch/software/fastsimcoal2/) [160]. With genome-wide data from a population level sample, produced either through a reduced genome representation technique (see Section 4.3) or resequencing (sequencing of a genome of using less coverage and a template draft genome sequence) population structure and demographic history can easily be estimated through these variety of approaches discussed above to provide powerful insights into the source and origins of agricultural weed populations.

While genome wide information provides high resolution data on the distribution of allelic differences among samples due to demography, allelic differences due to selection can be inferred with care using integrative summary statistics and model based approaches. Currently, our understanding is that HR evolution often proceeds through drastic changes in allele frequency at the target gene—conveniently, a single locus of large effect provides the most power for detecting recent signals of selection and differentiating independent events. Three types of signal can be used to

recognize selection: changes in allele frequencies (differentiation and diversity), patterns associated with linkage (homozygosity), and the pattern of nucleotide substitutions

First, regions near alleles selected by agricultural practices can be indicated by changes in allele frequencies. When a beneficial allele changes in frequency, becoming highly prevalent or fixed in a population sites nearby, linked to the selected allele due to a low probability of recombination, will show a depletion of genetic variation. The pattern resulting from the fixation of nearby neutral sites along with the selected site is termed a selective sweep [161–164]. An expectation following from this process is that the frequency of alleles under selection is expected to differ among populations experiencing different conditions (e.g., herbicide application or none) and this differentiation between populations is frequently expressed as Wright's fixation index (F_{ST}), though there are a host of related statistics [165,166]. If the F_{ST} of a locus is much larger than at other nearby or neutral loci, this can indicate positive selection.

Second, in addition to differentiation, immediately following selection the frequency of linked alleles will be fixed with new mutations causing new alleles to accrue slowly thereafter. This results in an excess of homozygosity (lack of variant sites) directly after selection. As new alleles will be rare, an excess of rare alleles can indicate positive selection (as well as recent population expansion) and can be quantified by Tajima's D, which compares the number of pair-wise differences between individuals with the total number of segregating polymorphisms [167]. Similarly, Fay and Wu compare the number of pair-wise differences between individuals to the number of individuals that are homozygous for the allele [168].

Third, selection can be detected through a comparison of the rate of nonsynonymous substitutions at a nucleotide (those that alter the amino-acid represented by the codon) to the rate of synonymous substitutions, which are assumed to be silent and neutral. This ratio can indicate selection favoring a change in the structure of a protein (d_N/d_S).

Beyond these summary statistics, many model-based approaches have been developed to distinguish between recent, single genetic origin selective events (hard sweeps) and older or multiple genetic origin selective events (soft sweep) by assessing differences in the magnitude of their signals across the genome (e.g., SweeD (cme.h-its.org/exelixis/web/software/sweed/index.html) [169] and SweepFinder2 (www.personal.psu.edu/mxd60/sf2.html) [170,171]). After assaying within population sweep patterns, one can then compare the extent of convergence in these patterns across populations. A greater or lesser extent of parallel changes in allele frequencies, homozygosity, and diversity in the surrounding sequence provide evidence of shared or independent origins of resistance across populations respectively, and more broadly, may provide the means to identify candidate genes that appear to underlie HR in multiple populations (see Section 4).

While there is great potential to determine the source and number of independent and shared origins of HR from genomic data (e.g., [88]), the task will be more difficult when HR is conferred by many alleles of small effect. With polygenic trait architectures many individuals are needed to have sufficient power to detect the individual small-effect changes, and therefore approaches often rely on taking the sum of allele frequencies weighted by their effect size on the trait [172]. Since these genome-wide association approaches assume allele frequency differences across the genome are all related to selection, one must carefully account for allele frequency changes due to population structure, which has been shown to often be confounded with polygenic signals of selection [173].

3.2. Example: Convergent Adaptation to Glyphosate in Common Waterhemp

Common waterhemp is a problematic, a wind-pollinated, outcrossing, and dioecious weed that occurs throughout the mid-western and eastern United States of America and in Canada from Manitoba to Quebec. It has been hypothesized that weedy agriculture populations result from human-mediated disturbance and mixing of two closely related taxa, *A. tuberculatus* var. *rudis*, a Midwestern native, highly associated with agricultural environments, and *A. tuberculatus* var. *tuberculatus*, a species that occupies a constrained range, and that is limited to riparian environments [174]. Glyphosate resistance

was first reported in 2005 in Missouri and one hypothesis is that it may have spread from there across the United States and recently into Ontario. However, considering the strength of selection from herbicides and the highly repetitive nature of HR evolution as suggested from independent glyphosate resistance evolution in multiple *Amaranthus* species [35], it is also possible that glyphosate resistance may have multiple independent origins with *A. tuberculatus*, representing a striking case of convergent evolution.

A recent study used genomic approaches to investigate the history of the species, clarify the origins of agricultural populations, and the evolution of glyphosate resistance [88]. Specifically, Kreiner et al. [88] sequenced the species' genome as described above (see Section 2.6) and then resequenced the genomes of 163 individuals from 19 agricultural populations known to have glyphosate resistance, varying from 13% to 88% of the population, from Missouri, Illinois, and Essex County and Walpole Island within Ontario, as well as ten individuals from a native, non-agricultural population in Ontario that lacked glyphosate resistance. This data and the software freebayes (github.com/ekg/freebayes) [174] were used to identify SNPs across the species genome and then to characterize population demographics, diversity, differentiation, and structure. Demographic modeling completed using $\delta\text{a}\delta\text{i}$ supported the hypothesis of recent secondary contact between lineages. Similarly, analysis with STRUCTURE and principal component analysis, indicated that populations were genetically differentiated by geography and hypothesized species ranges, with populations from Missouri and Illinois clustering and corresponding to *A. tuberculatus* var. *rudis* and natural populations from Ontario clustering and corresponding to *A. tuberculatus* var. *tuberculatus*. These analyses also showed resistant populations from Essex county were unlike nearby natural or agricultural populations found in Ontario, but rather clustered with western Missouri populations. This indicates that populations from Essex County likely represent an introduction of seed from Midwestern *A. tuberculatus* var. *rudis* populations, that harbored multiple independent resistance haplotypes. Interestingly, the second group of resistant populations in Ontario, those from Walpole Island, clustered with natural populations in the area, though with signs of some introgression from the var. *rudis* cluster. With information on the evolutionary origins of these populations, Kreiner et al. set out to distinguish whether populations with shared evolutionary origins have independently evolved resistance, or if resistance spread through the expansion of these populations into new agricultural landscapes. The authors investigated the pattern of selection on the chromosome bearing the glyphosate target-site gene, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), using Sweepfinder2 and model-free summary statistics such as diversity, homozygosity, and differentiation. This analysis indicated the plants from Walpole showed a stronger pattern of reduced genetic diversity, increased differentiation and increased extended haplotype homozygosity around the EPSPS genes—evidence of a hard selective sweep—distinct from plants from Essex county, Missouri, or Illinois where a soft-sweep following multiple origins throughout the Midwest appears to have occurred. The authors conclude that glyphosate resistance in newly problematic Ontario populations has multiple genetic origins – both through new seed introduction events and selection on a recently arisen mutation in a previously benign population.

4. Current Application: What Genes Underlie Herbicide Resistance?

Understanding the genetic basis of resistance to an herbicide in a plant species is an essential first step in the development of diagnostic markers, understanding the fitness consequences of the mutation, and, more generally, in understanding how herbicide evolution typically occurs. This information is essential for being able to detect, monitor and develop more effective strategies for managing HR. Of the current total of 500 unique combinations of species (256) and herbicide site mode of action, the underlying genetic basis of these resistances is only known for a minority of cases [35]. The majority of known cases involve mutations to the herbicide's target site (TSR), while the specific genetic basis of non-target site resistance (NTSR) is largely unknown [175,176].

Our lack of understanding of the genetic basis of NTSR, is a major gap in our understanding of weed biology and the evolution and spread of HR [175–177]. Non-target site resistance is the most

common mechanism contributing to glyphosate and acetyl CoA carboxylase inhibition resistance (ACCase). It is also the most common mechanism for acetolactate synthase (ALS) resistance in grass species [178] and can confer resistance to several herbicide modes of action simultaneously and unpredictably [179]. Non-target site resistance encompasses a diverse and complex set of traits that likely involve the full gamete of potential genetic basis including dominant to semi-dominant alleles with major effects, copy number variation, multiple minor alleles that incrementally contribute to resistance, and changes in epigenetic regulation (reviewed by [180]). Further, NTSR likely involves varied aspects of the fundamental processes within cells from transcription to translation invoking complex stress responses and altering regulatory pathways [177,180,181]. Until this gap in our knowledge is filled in our ability to make diagnostic tests, draw conclusions about the type and prevalence of mutations/variation that contribute to HR or develop strategies to interfere with NTSR pathways is compromised. However, while we rarely know the specific genetic basis of NTSR in a weed species, we have a good understanding of the types of genes are most likely involved.

4.1. Five Superfamilies of Suspects

Five gene superfamilies have members that have been identified as likely involved in NTSR. Evidence for their involvement comes from either their ability to confer herbicide tolerance or resistance in crop species or *Arabidopsis*, on enzyme and transcriptome analyses of herbicide resistant species or investigations of the molecular mechanisms of drug resistance (reviewed by [182,183]). Evidence from transcriptome studies suggests NTSR is often the result of the action of multiple members of a superfamily and multiple superfamilies [184–187]. Each of these families are large, diverse, and widely represented across the tree of life from bacteria to mammals indicating that they are fundamental to how organisms cope with their environments. In this regard, the evolution of HR has selected variants of genes underlying the complex regulatory and enzymatic pathways that organisms have always used to face biotic and abiotic stresses [188]. These gene superfamilies are considered to form part of what has been termed the xenome, the chemical detection, transport and detoxification system of plants [189] and members of the families are spread throughout plant genomes.

4.1.1. Cytochrome P450 Monooxygenases

The cytochrome P450 monooxygenase gene superfamily (CYP) are the largest enzyme family in plants and are known to be involved in HR [190]. This superfamily, which is involved in detoxification and stress responses, were implicated in HR as a result of the analysis of herbicide residues from plants, their induction following the application of safeners (chemicals that increase herbicide tolerance in grain crops), and the observation of increased P450 metabolism levels in HR annual ryegrass (*Lolium rigidium* Gaud.), black grass (*Alopecurus myosuroides* Huds.) and lesser canary grass (*Phalaris minor* Retz.) [177]. However, the number of these genes [191], with 272 in *Arabidopsis thaliana*, for example [192], and issues with purification from plant material meant that the isolation of specific CYP genes conferring HR in plants was preceded by isolation of these genes in bacteria and mammals, which frequently have higher activity than those from plants [193]. As an example, expression of human CYP genes in potato [194] and rice [195–197] confer HR. Indeed, expression of *CYP1A1* in rice resulted in resistance to ten different herbicides from ten different Herbicide Resistance Action Committee (HRAC) groups [195,198], while expression of *CYP2B6* in resistance to thirteen from six HRAC groups [197]. Despite this demonstrated ability of individual CYP genes to confer broad HR, it is likely that multiple CYP genes are involved in NSTR within each plant species [177]. Plant derived CYP genes that have been demonstrated to confer HR have now been isolated in Jerusalem artichoke (*Helianthus tuberosus* L.) [199], soybean (*Glycine max* (L.) Merr.) [200], *Arabidopsis* [201] and ginseng (*Panax ginseng* Mey.) [202]. Within weeds, two CYP genes have been determined to be associated with ALS resistance in rice barnyardgrass (*Echinochloa phylllopogon* (Staf.) Koso-Pol.) and overexpression of these genes in *Arabidopsis* resulted in resistance to group B herbicides bensulfuron-methyl and penoxsulam [203] and group F4 clomazone [204]. The isolation of CYP genes responsible for HR from other weed species will

likely occur in the near future as chemical inhibition of P450 indicate that these genes are involved in HR for flixweed (*Descurainia sophia* L.) [205], water hemp (*Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer) Costea & Tardif [206], and large crabgrass (*Digitaria sanguinalis* L. Scop.) [207] in addition to the grass species mentioned above. Additionally, consistent expansion of CYP copy number across all 69 annotated CYP genes in *Amaranthus tuberculatus* agricultural populations relative to natural populations has been recently found [88].

4.1.2. Glutathione S-Transferases

Glutathione S-transferases (GSTs) are enzymes that play a strong role in plant secondary metabolism and stress response [208–210]. For example, GSTs have been identified as playing a role in salt tolerance [182], copper tolerance [211] and fungal disease resistance [212]. They were first identified in mammals in the 1960s because of their role in drug metabolism and their presence in plants was identified soon after as contributing to atrazine resistance in maize (*Zea mays* L.) [213]. As a result, the role of GSTs for herbicide detoxification in maize have been extensively studied [214] and several of the genes encoding these enzymes have been used to engineer HR. For example, *GST1* [215] expressed in tobacco (*Nicotiana tabacum* L.) [216], resulted in resistance to alachor (group K3) and *GST27*, when expressed in wheat (*Triticum aestivum* L.), resulted in atrazine (group C1) and oxyfluorfen (group E) resistance [217]. Similarly, overexpression of a GSTs from soybean, *GmGSTU4*, in tobacco results in a significant increase in alachor tolerance [218]. Within weeds, two glutathione-S-transferase genes have been identified as being involved in resistance to ACCase and ALS inhibitors in black grass [219]. Indeed, although multiple loci are believed to be involved in NTSR HR for black grass [220], expression of *AmGSTF1* in *Arabidopsis* resulted in resistance to atrazine, alachor, and chlorotoluron (group C2) [185]. Expression analysis suggests that GSTs are involved in HR for a number of other weed species including junglerice (*Echinochloa colona* (L.) Link.) [221], Palmer amaranth (*Amaranthus palmeri* S. Wats.) [222], annual ryegrass [184,223] and sunflower (*Helianthus annuus* L.) [224]. However, as with the CYP genes, the number of GSTs in a plants species makes pinpointing the specific gene or genes responsible for HR challenging. For example, there may be 42 in maize [225] and 54 functional GSTs have been identified in *Arabidopsis* [226].

4.1.3. ATP-Binding Cassette Transporters

ATP-binding cassette (ABC) transporters are a group of proteins that mediate cross membrane transport (reviewed by [227,228]). With more than 80 members they are the largest protein family in *Escherichia coli*. Approximately 130 and 150 members have been located within the *Arabidopsis* [229] and the tomato (*Solanum lycopersicum* (L.) H. Karst.) [230] genomes, respectively. These transporters are understood to be involved in the transport of auxin and glyphosate and may, therefore, play a role when reduced translocation or sequestration of these herbicides is involved in HR [177,231]. In horseweed (*Conzuya canadensis* (L.) Cronq.) glyphosate application caused increased expression level in at least seven ABC transporter genes [232] and a transcriptome study on the closely related hairy fleabane (*Conzuya bonariensis* (L.) Cronq.) indicated that there were 19 ABC transporter genes in addition to 22 other candidates including GSTs and glycotransferases (see below). Additional evidence of the role of this group is that overexpression of the ABC transporter gene *AtPgp1* in *Arabidopsis* resulted in resistance to dicamba (group O) and oryzalin (group K1) [233] and tobacco overexpressing *pqrA* from the bacterium *Ochrobactrum anthropi* show higher resistance to paraquat (group D) [234].

4.1.4. MFS Transporters

The major facilitator superfamily (MFS) are also transporter proteins. As with the ABC transporters, there are approximately 70 members of the family within the genome of *Escherichia coli* [235] with perhaps 200 in *Arabidopsis* [236]. Like the ABC transporters members of the MFS family have been identified as being upregulated following exposure to auxinic herbicides [237] and the *TPO1* gene from yeast is a member of this group and its homolog from *Arabidopsis*, *At5g13750*, are able to confer

resistance to 2,4-D when overexpressed in yeast [238]. However, it does not appear that studies examining the consequences of over expression of this type of gene in plants have been completed.

4.1.5. Glycosyltransferases

Glycosyltransferases (GTs), enzymes that add carbohydrates to molecules, are involved in the detoxification of herbicides in addition to many other roles within plant cells [239,240]. They are numerous in plant genomes with one particular family within this superfamily, the UDP-glucose dependent glycosyltransferases (UGTs), having 107 functional members in *Arabidopsis* [241]. Like CYP and GSTs genes, they are induced by the application of safeners and have been detected in transcriptome studies following herbicide application [189] and enzymes from this group from a wide variety of organisms have been demonstrated to have activity against atrazine and fluorodifen (group F1) [240]. However, unlike the other superfamilies discussed here, we did not find any examples of genes from this family being used to produce HR organisms. Instead, much of the work focused on these enzymes is examining the potential of these enzymes in phytoremediation of organic pollutants [189,242,243]. For example, a gene in *Arabidopsis* (*UGT72B1*) encodes an enzyme that detoxifies 3,4-dichloroaniline (DCA) and 2,4,5-trichlorophenol (TCP) [244].

4.2. A Role for Genomic Approaches

Due to the complexity, diversity, and number of genes that could underlie NTSR; identification of resistance-conferring mutations is a significant challenge even when one has a lead on the potential genetic basis from the above insights [180]. Clearly, significant progress is being made through the application of RNA sequencing to identify the genes being expressed following herbicide application, expression analysis of those genes using quantitative PCR, and transformation of model organisms such as *Arabidopsis* and tobacco to verify the function of the genes. Additional genomic information for weeds is an asset for this type of investigation and can allow comparative genetic approaches and searches with tools such as BLAST [106] to identify and classify members of the multigene families discussed above as has been done in model organisms and crops (e.g., [225]). This can allow for systematic testing of the activity each enzyme (e.g., [241,245]). However, there are undoubtedly more genes and gene families involved in NTSR (e.g., [246]). As with unravelling the demographic history and structure of populations discussed, one method of identifying these genes is to examine the signature of the strong artificial selection pressure of herbicide application across the genome (see Section 3.2). Additionally, a physical map combined with the tools of genetics (e.g., linkage mapping, genome-wide association studies) can inform on small to large effect genomic loci involved in HR.

4.3. Example: Glyphosate NTSR in Morning Glory

A recent *tour de force* investigating glyphosate resistance in morning glory (*Ipomoea purpurea* (L.) Roth.) provides a clear example of how genomics and detection of the signature of selection can be applied to understanding the basis of non-target site resistance. In this work, Van Etten and colleagues [187] generated genome wide DNA markers to examine population structure, the possibility of multiple origins of HR in the species, and to provide an indication of where selection was acting in the genomes. They then sequenced the species' genome and re-sequenced targets within the exome, the regions of the genome that are the parts of a gene that encode the final RNA transcripts, in regions showing selection. This data was used to assemble multiple lines of evidence to identify the candidate genes underlying glyphosate resistance.

To provide information of population differentiation and structure, examine the evidence for HR genes being introduced to populations via gene flow versus the HR arising multiple times, and to search for signatures of selection Van Etten et al. [185] used a reduced genome representation technique (nextRAD). This approach identified single nucleotide polymorphisms (SNPs) across the species' genome for ten individuals from each of four high and four low survival populations. This approach is a variant of restriction site associated DNA sequencing (RADseq), which in general, use restriction

enzymes (often a pair) to selectively amplify regions adjacent to restriction sites across a species' genome [247,248]. The number of markers can be manipulated through the length of the restriction enzyme's recognition site allowing for the density of the markers to be manipulated depending on the project's goal and species genome size. As no sequence data is required before hand, this type of data can be generated for species whether or not they have genome sequences available. For each individual, enough Illumina sequencing data needs to be completed to result in approximately 30× coverage for each amplified region. Then programs such as STACKS (catchenlab.life.illinois.edu/stacks) [249–251] or TASSEL (bitbucket.org/tasseladmin/tassel-5-source/wiki/Home) [252] can be used to either group reads by similarity, if a sequenced genome is unavailable, or to align the reads to a draft genome sequence to locate polymorphic (variable) SNPs. These SNPs can then be analyzed with a plethora of packages in the free statistical programming language R [253] to understand the population biology (reviewed by [254]). This can include calculation of population differentiation (F_{ST}) using hierfstat [255] or StAMPP [256]; the generation and visualization of unweighted pair group method with arithmetic mean (UPGMA) or neighbor joining trees using poppr [257] and phytools [258]; and k-means clustering (adeget [259]) to further investigate population structure. In the case of glyphosate resistance, in both morning glory [185] and Palmer amaranth (*Amaranthus palmeri* S. Wats.) [260], this approach indicated that gene flow introducing HR alleles has likely been responsible for much of the pattern of resistance and susceptible populations. However, in addition to gene flow, a second origin of glyphosate resistance was also suggested in Palmer amaranth [260].

The population level SNP data generated by Van Etten et al. [185] was then further analyzed with two programs, BayeScan [261], which can identify SNPs that show signs of selection and bayenv2 [262], which indicate SNPs associated with levels of HR. BayeScan (cmpg.unibe.ch/software/BayeScan/) calculates pairwise F_{ST} values between each population sampled and a theoretical population comprised of a common gene pool from all sampled populations. Selection is implied as an explanation, if a locus specific factor improves the logistic regression model for these F_{ST} values that includes population structure [261]. The program bayenv2 (bitbucket.org/tguenther/bayenv2_public/src/default/) looks for correlations between an environmental variable, such as HR level, and SNP frequency using a Bayesian method that estimates the pattern of covariance of allele frequencies, uses this as a null model and then tests each SNP [262]. Putative genes in proximity to the 42 outlier SNPs identified by BayeScan and the 83 SNPs flagged by bayenv2 were then identified by annotation tools such as AUGUSTUS (see above).

Next they sequenced a morning glory (diploid, approximately 978 Mb, 1C = 1.0 pg [24], 2n = 30 [24]) individual that they considered to be high homozygous using PacBio reads (11 SMRT Cells) and Illumina short read data (100 bp paired end). They completed two genome assemblies one using only the Illumina data with the program ABYSS (github.com/bcgsc/abyss) [263] and the other using a hybrid approach that combined their long and short read data with the program DBG2OLC (github.com/yechengxi/DBG2OLC) [264]. This later assembly consisted of 17,897 scaffolds, had an N50 of 15,425 and a total length of 1,948 Mbp.

They then used their genome assembly to design probes (baits) to perform target-capture resequencing of these genes, the EPSPS genes, genes previously associated with HR and a randomly selected control group. This targeted exome re-sequencing was then completed for five individuals from each of their eight populations. These re-sequenced contigs were aligned to the chromosome level sequence of Japanese morning glory (*Ipomoea nil* (L.) Roth.) [265] to visualize the pattern of outliers indicating selection and they identified five regions of interest which contained 945 genes—including multiple members of the CYP, GSTs GT, and ABC transporter superfamilies. To determine if the number of members identified in these regions was greater than expectation for these large families, they resampled Japanese morning glory's genome to provide a baseline estimate of the number of that would be expected. This indicated that GT, ABC transporters and CYP genes were each overrepresented in the identified regions. These five regions also showed high genetic differentiation between populations with high and low glyphosate survival.

One approximately 29 kb region aligned to Japanese morning glory's chromosome 10 showed reduced nucleotide diversity in resistant individuals, strong evidence of selection based on Tajima's D and Fay and Wu's H as well as stronger linkage among the SNPs of this region. This region contained a tandemly repeated group of seven GT genes and nine CYP genes. For this region, they determined that the majority of resistant individuals shared high genetic similarity and tests of convergence suggesting that this region contains one or more beneficial genes that were introduced by gene flow and rapidly swept through resistant populations. While none of the non-synonymous SNPs in these genes showed fixation in the high survival populations, this region has a strong likelihood of containing loci that underlie glyphosate resistance in the species and are strong candidates for further functional validation.

5. Future Application: Can We Genetically Alter Weed Population to Make Them Easier to Control?

With a greater understanding of the population biology of weed species and the identification of the DNA sequence changes that underlie HR come opportunities for new control strategies. This is made particularly true by the development of genetic engineering methods involving clustered regularly interspaced short palindromic repeats (CRISPR) technologies. CRISPR tools are both simple and versatile, contributing to their successful spread in all aspects of molecular biology (reviewed in [266]). CRISPR systems are found in bacteria and archaea where they provide acquired immunity against invasive elements like phages. They do so by co-opting small pieces of DNA sequence from the pathogen which they subsequently use to generate guide RNA molecules that "program" an endonuclease (e.g., Cas9) to scan the genome and find its target. The recognition of DNA sequence homologous to the guide triggers the cleavage of the DNA strand that leads to mutations and potential inactivation of the targeted element.

In a landmark study, the CRISPR system of *Streptococcus pyogenes* was reduced to two components, an endonuclease (Cas9) and a single guide RNA, that could efficiently and specifically cut DNA in vitro [267]. Following this, similar two-component systems were introduced in a plethora of different organisms to engineer mutations in the DNA sequence with outstanding success [268]. Ultimately, the only requirement for this approach is the knowledge of the targeted DNA sequence, making application in weed control theoretically possible [269]. Consequently, while the short answer to the question "Can we genetically alter weed population to make them easier to control?" is *probably*, there are a great number of technical [270], ethical [271] and ecological [272] hurdles and no current examples of this approach being used in weed science. Here we focus on describing and discussing the potential and technical challenges to developing a weed control strategy using the engineering of whole populations. For an example, we reach beyond weed science to the control of insecticide resistant mosquitoes, summarizing the current findings and approaches of the scientists, who are likely to be the first to release gene drive element into the environment to control a pest population.

5.1. The Potential for Manipulation of Weed Populations

Ever since the demonstration of the repurposing of a bacterial CRISPR system as a programmable endonuclease [267], there has been speculation about its potential use for pest control or eradication [273]. Indeed, there were early successes in the application of CRISPR-based "gene drive" systems in order to decimate or modify populations of fruitfly (*Drosophila melanogaster* Meigen) and importantly, disease-spreading mosquitoes (*Anopheles stephensi* Liston and *Anopheles gambiae* Giles) (reviewed by [274]). The basis of a gene drive system relies on using a selfish genetic element capable of either copying itself or biasing reproduction towards its own inheritance so that it propagates through a population in a non-Mendelian fashion. This cheating of the classic inheritance rules can compensate for some deleterious consequences and potentially allow a measure of population control. Adding CRISPR components to this paradigm then allowed homing in on specific targets within the genomes making it available to newly sequenced weed plants [269]. Such a system has yet to be created in plants, but the rapid evolution of plant genetic engineering could make it a reality in the not too distant future.

Indeed, in their report “Gene Drives on the Horizon” the National Academy of Sciences considers the potential of this strategy for the control of Palmer amaranth [271].

The overarching goal of such an endeavor is to create a transgenic weed able to introduce a genetic payload into the populations of its species using biased inheritance and resulted in populations that are easier to control because of a vulnerability introduced with the payload. What the ideal payload would be up for debate, but it is likely to include a CRISPR system composed of a gene encoding a programmable endonuclease like the *Streptococcus pyogenes* Cas9 and a single or multiple guide RNA. These guide RNA could be specifically designed to pair with the locus causing HR or, if this basis is unknown, the target locus could be unrelated to the HR allele, with the goal of introducing sensitivity to a new molecule altogether. The recognition of the target triggers catalytic activity and the cutting of the target DNA creating a lesion. Since DNA breaks are highly detrimental, they are quickly repaired by one of the many pathways existing in the host cell. The gene drive system then subverts the DNA repair pathways ensuring its own propagation. This step represents one of the major challenges to this approach, as plant cells are known to heavily favor non-homologous DNA repair pathways that only produce small DNA sequence changes [275] that would fail to propagate the selfish element.

Indeed, the success of gene drive methods in fruitflies and mosquitoes is due in large part to the frequent use of homology-guided DNA repair in insect cells. However, plant somatic cells seldom use homologous recombination and favor non-homologous repair mechanisms [275]. For gene drive elements to spread efficiently in a plant population, this ratio between the two types of repair would have to be altered. This would be critical as non-homologous repair would create alleles resistant to the CRISPR system that would counter efforts to spread the gene drive. This is why the precise insertion of the gene drive element at a chosen location in the weed genome will likely be a *sine qua non* condition to its propagation. Once integrated, the new allele can start competing with natural alleles, which it can target for cleavage and convert using the host cell machinery. Encouragingly, the molecular mechanism called gene targeting, which uses the same homologous host DNA repair pathways as the gene drive approach, is of great interest in plant genetic engineering and has greatly improved the past few years [276]. Gene targeting aims at delivering a DNA sequence of interest at a specific location within the genome and, therefore, has also greatly benefited from advances in CRISPR technologies. Just like gene drive, gene targeting requires the use of homology-guided DNA repair mechanisms instead of non-homologous DNA repair. The difference between the two is that the final goal of gene targeting is a single isolated event, while a gene drive must self-propagate indefinitely, thereby adding to the challenge.

Excitingly, the case of a bacterial transposon that co-opted a CRISPR system as a means to guide its own propagation within the genome was recently discovered [277]. Transposons are themselves selfish elements that have evolved different means to copy themselves to favor their propagation. For example, some transposons encode an enzyme called integrase that can insert a DNA fragment at a target site in a genome. This new molecular tool has enormous potential as a gene drive system being able to circumvent the need to coax the host repair machinery to use homologous repair mechanisms.

5.2. Additional Technical Challenges

There are a number of additional technical limitations in the creation of a useful gene drive system for weed management beyond a need for the target species to use homologous repair mechanisms. As a first hurdle, this approach would be restricted to plants that can be genetically transformed and little effort has been devoted to the development of transformation techniques in weeds. Plant susceptibility to transformation is highly variable and whether or not it is ultimately possible in a species depends on many intrinsic factors [278]. For instance, species with unfused carpels at the extremity of the stigma may be amenable to the convenient floral dip *Agrobacterium* mediated transformation method. However, the great majority of plant species relies on other methods, such as tissue culture with *Agrobacterium tumefaciens* Smith and Townsend or biolistic bombardment, both being much more time and resource consuming. It could, therefore, take a few months to many years

to develop a new transformation protocol for a particular plant—a potentially sizable initial investment of resources.

When transformation is possible, the challenge of precisely integrating a given DNA construct remains. At the molecular level, the problem can be broken down into two distinct parts; the mobilization of the homologous repair machinery and the delivery of the DNA template to be copied in the genome. For the first part, it has been reported that expressing the CRISPR system in specialized cells where homology-guided DNA repair occur at higher frequencies can increase gene targeting [276]. We know for instance that cells undergoing meiosis rely on homologous recombination between DNA molecules for orchestrating proper chromosome segregation. One could take advantage of these cell-specific conditions and engineer a system that would only act in a specific cell context as was recently done in mouse female germline [279]. Another interesting avenue is the tethering of repair machinery components to the endonuclease. Indeed, the fusion of Cas9 with different proteins offers many opportunities including influencing downstream DNA repair as it was successfully done in human cells [280]. Such an approach could be tailored to improve the propagation success of a gene drive element. In the second part of the molecular cascade, a DNA template has to be provided for the homologous repair machinery to integrate at the break site. In the case of gene drive, the engineered allele would bear homology to the wild allele and would therefore present itself as a repair template. Interestingly, recent studies have shown increased success in gene targeting when using components of a geminivirus [281–283]. The rationale behind this approach is that viruses can generate multiple extrachromosomal copies of a given DNA sequence thereby increasing the chances of any one fragment being used as template by the repair machinery. This element could be included into a gene drive system to increase its efficiency.

5.3. Evolutionary Consequences and the Need for Integration with Other Management Strategies

Even without the numerous technical impediments to gene drive strategies in weeds, this approach presents enormous ethical, regulatory, and ecological challenges. Theoretically, a gene drive that reduces the fitness of a population or its ability to reproduce could bring a species to extinction, as it was convincingly demonstrated for caged mosquitoes [284]. Setting this as a goal seems unwise and unlikely to gain societal support [272,285,286] or regulatory approval [287], as a result, strategies to re-sensitized populations to an herbicide or create susceptible to a specific compound unlikely to be found beyond the agroecosystem are likely to be more tenable. The advantage of such an approach is that it does not reduce the fitness of the population in the wild per se. Like the use of herbicides, altering weed populations as a management strategy would not be a silver bullet and would require integration into integrative weed management strategies. In part, this would be a consequence of the time needed for alleles to spread through populations as this could take 10 to 20 generations for a gene drive system to saturate a population [288]. In the re-sensitizing approach, this would mean forsaking the use of a given herbicide for many years thereby relying on other control strategies. In this regard, creating a susceptibility to a new molecule would present advantages but great care would need to be taken in choosing such a compound.

A second reason why this strategy would need to be part of an integrated weed management strategy, comes from the lesson we have learned from our reliance on herbicides. Plants are quite able to evolve in response to selection through modification of genetic machinery, the exome (see Section 4), and the biotic challenge represented by a gene drive element will result in selection on similar genetic machinery used to counter similar genetic attacks from viruses or selfish genetic elements. For example, in the case of a CRISPR-based gene drive, any synonymous mutation to the targeted site(s) would severely reduce the efficiency of the endonucleolytic cleavage [289]. This has already been demonstrated in model species such as fruitflies [290]. The emergence of such allele would be expected and could be mitigated by selecting sites where mutation would have high fitness cost would be more likely to provide a robust solution [291]. Since CRISPR genes come from bacteria, there is also a chance the plant cell would silence them using intrinsic mechanisms and a silenced allele could

render then organism “immune” to the subsequent use of a CRISPR-based approach. Taken together, all these considerations argue for thorough modelling and confined population studies before such a strategy could be released in the fields as has been laid out in recommendations by the National Academy of Sciences [271].

5.4. Example: Gene Drive in Malaria Vector Mosquitos

While examples of gene drive development in weed species remain for future reviews, significant work has focused on using the technology to control mosquitoes that spread malaria. This is a system with parallel challenges to those faced in weed science including the emergence of multiple-insecticide resistance with both target site and NTSR mechanisms and a lack of new chemical control options [292,293]. Malaria is a serious and prevalent disease with over 200 million cases a year. It is often fatal, particularly in children, and disproportionately affects people living in South America, South Asia and sub-Saharan Africa where access to health care is often limited. The World Health Organization reported that of the 435,000 deaths reported in 2017 from malaria, ninety-two percent occurred in Africa and sixty-two percent occurred in children under five [294]. Malaria can be caused by any one of five *Plasmodium* parasites and can be transferred by several of the 450 species of *Anopheles* mosquitoes [294]. Within the sub-Saharan Africa region, malaria is primarily the result of infection by *Plasmodium falciparum* Welch transferred by female *Anopheles gambiae* mosquitoes [294]. Chemical strategies for controlling populations of these mosquitoes have resulted in the evolution of insecticide resistance with the first cases of pyrethroid resistance reported in Sudan in the 1970s and reports of resistance now available across Africa and in Madagascar [295]. Currently, *A. gambiae* populations in regions such as the Côte d’Ivoire and Burkina Faso, have evolved complete resistance to all approved classes of insecticides [296,297]. In 2015, researchers developed a CRISPR-based gene drive system designed to reduce reproductive capability by disrupting the sequence of a gene likely involved in the development of the embryo’s body plan which results in female sterility. When carriers of this this gene were crossed to wild type mosquitoes the gene had a transmission rate of just over 99% and it was able to spread through a caged populations initiated from equal numbers of wild type and transformed individuals [298]. However, nuclease-resistant variants that completely blocked the spread of the gene could be detected as early as the second generation [285]. More recently, in 2018, the researchers improved on these results by disruption of a gene that controls sex differentiation and that has alternative splicing patterns in male and female mosquitoes, a characteristic believed to increase the constraints in the development of resistant variants. One of the two cages, initiated with 12.5% disrupted allele frequency, reached 100% allele frequency at generation 7 and extinction at generation 8, while for the second cage these two points were reached at generation 11 and 12 respectively. Importantly, they did not detect an evidence for the evolution of resistance to this gene drive, though they note that it may not be “resistance-proof” given a wider sample of mutations [284]. This work relied on foundational genomic information from *A. gambiae*’s genome sequence in 2002 [299] as well as detailed knowledge of the genetic basis of fundamental aspects of *A. gambiae*’s biology. In July 2019, the researchers initiated small scale releases of genetically modified, sterile males (not equipped with gene drive) in Burkina Faso to produce the data required to meet the ultimate goal of releasing individuals with gene drive to control malaria [300]. The researchers that have developed this technology work with a consortium, Target Malaria (targetmalaria.org), that includes scientists, regulators, and community engagement specialists. They have also worked to understand the ecological risks associated with the unconfined release of this event [301]. This approach to develop the social license and regulatory approval for this type of intervention provides a valuable template for how weed scientists could approach the modification of a weed species for population management.

6. Conclusions

Genomic approaches are extremely powerful tools for understanding biological systems. These tools, while currently underutilized in weed biology, are exciting in their potential to answer key weed science questions and increasingly accessible. Here our goal is to provide a foothold for weed scientists considering this type of research by providing an introduction to the considerations and process of creating a draft genome and illustrating how that genome could be used as a fundamental tool. Draft weed genomes can provide a resource for demographic analyses that examine the result of selection on the genome. This information can shed light on the evolutionary origins of weeds allowing us to identify management practices that could prevent HR evolution. It can identify strengths and weaknesses of weed populations that can be targeted for control, while providing fundamental information on how plants rapidly respond to selection from humans. The changes that selection makes to the genome and revealed by genomic approaches can also provide evidence of which loci are the genetic basis of NTSR. This information will allow us to form strategies to interfere with these HR mechanisms. Finally, the insights we gain from a better understanding of weed species at the population, genomic and genic level using these approaches open the option of altering the genome of weed species to provide us another tool for weed management—a strategy nearing implementation in mice and mosquitoes.

Author Contributions: All authors contributed to the conceptualization, writing and editing of this paper. Funding acquisition was co-led by E.P and M.L.

Funding: This research was funded by Agriculture and Agri-Food Canada (AAFC), “Deciphering complex mechanisms and inheritance patterns of herbicide resistance cases in Canada” grant number J-001751.

Acknowledgments: We thank Tyler Smith, Connie A. Sauder and Beatriz E. Lujan-Toro for comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666–681. [[CrossRef](#)] [[PubMed](#)]
2. Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1–8. [[CrossRef](#)] [[PubMed](#)]
3. Li, F.-W.; Harkess, A. A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **2018**, *6*, 1–7. [[CrossRef](#)] [[PubMed](#)]
4. Dominguez Del Angel, V.; Hjerde, E.; Sterck, L.; Capella-Gutierrez, S.; Notredame, C.; Vinnere Pettersson, O.; Amselem, J.; Bouri, L.; Bocs, S.; Klopp, C.; et al. Ten steps to get started in Genome Assembly and Annotation. *F1000Research* **2018**, *7*, 148. [[CrossRef](#)]
5. Armstrong, O.; Fiddes, I.T.; Diekhans, M.; Paten, B. Whole-Genome Alignment and Comparative Annotation. *Annu. Rev. Anim. Biosci.* **2019**, *7*, 41–64. [[CrossRef](#)]
6. Gillings, M.R.; Paulsen, I.T.; Tetu, S.G. Genomics and the evolution of antibiotic resistance. *Ann. N. Y. Acad. Sci.* **2017**, *1388*, 92–107. [[CrossRef](#)]
7. Loman, N.J.; Pallen, M.J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **2015**, *13*, 787–794. [[CrossRef](#)]
8. Hatfull, G.F. Bacteriophage genomics. *Curr. Opin. Microbiol.* **2008**, *11*, 447–453. [[CrossRef](#)]
9. Holmes, E.C. Viral Evolution in the Genomic Age. *PLoS Biol.* **2007**, *5*, e278. [[CrossRef](#)]
10. Gudbjartsson, D.F.; Helgason, H.; Gudjonsson, S.A.; Zink, F.; Oddson, A.; Gylfason, A.; Besenbacher, S.; Magnusson, G.; Halldorsson, B.V.; Hjartarson, E.; et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **2015**, *47*, 435–444. [[CrossRef](#)]
11. Stranger, B.E.; Nica, A.C.; Forrest, M.S.; Dimas, A.; Bird, C.P.; Beazley, C.; Ingle, C.E.; Dunning, M.; Flicek, P.; Koller, D.; et al. Population genomics of human gene expression. *Nat. Genet.* **2007**, *39*, 1217–1224. [[CrossRef](#)] [[PubMed](#)]

12. Altshuler, D.L.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Collins, F.S.; De La Vega, F.M.; Donnelly, P.; Egholm, M.; et al. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467*, 1061–1073.
13. Li, J.Z.; Absher, D.M.; Tang, H.; Southwick, A.M.; Casto, A.M.; Ramachandran, S.; Cann, H.M.; Barsh, G.S.; Feldman, M.; Cavalli-Sforza, L.L.; et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **2008**, *319*, 1100–1104. [[CrossRef](#)] [[PubMed](#)]
14. Ravet, K.; Patterson, E.L.; Krähmer, H.; Hamouzová, K.; Fan, L.; Jasieniuk, M.; Lawton-Rauh, A.; Malone, J.M.; McElroy, J.S.; Merotto, A.; et al. The power and potential of genomics in weed biology and management. *Pest Manag. Sci.* **2018**, *74*, 2216–2225. [[CrossRef](#)] [[PubMed](#)]
15. Basu, C.; Halfhill, M.D.; Mueller, T.C.; Stewart, C.N. Weed genomics: New tools to understand weed biology. *Trends Plant Sci.* **2004**, *9*, 391–398. [[CrossRef](#)] [[PubMed](#)]
16. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [[CrossRef](#)] [[PubMed](#)]
17. The Arabidopsis Genome Initiative Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)]
18. Michael, T.P.; Jackson, S. The First 50 Plant Genomes. *Plant Genome* **2013**, *6*, 1–7. [[CrossRef](#)]
19. Veeckman, E.; Ruttink, T.; Vandepoele, K. Are We There Yet? Reliably Estimating the Completeness of Plant Genome Sequences. *Plant Cell* **2016**, *28*, 1759–1768. [[CrossRef](#)]
20. Jung, H.; Winefield, C.; Bombarely, A.; Prentis, P.; Waterhouse, P. Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. *Trends Plant Sci.* **2019**, *8*, 1–25. [[CrossRef](#)]
21. Ekblom, R.; Wolf, J.B.W. A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.* **2014**, *7*, 1026–1042. [[CrossRef](#)] [[PubMed](#)]
22. Wajid, B.; Serpedin, E. Do it yourself guide to genome assembly. *Brief. Funct. Genom.* **2016**, *15*, 1–9. [[CrossRef](#)] [[PubMed](#)]
23. Leitch, I.; Johnston, E.; Pellicer, J.; Hidalgo, O.; Bennett, M. Angiosperm DNA C-Values Database. Available online: <https://cvalues.science.kew.org/> (accessed on 28 May 2019).
24. Rice, A.; Glick, L.; Abadi, S.; Einhorn, M.; Kopelman, N.M.; Salman-Minkov, A.; Mayzel, J.; Chay, O.; Mayrose, I. The Chromosome Counts Database (CCDB)—A community resource of plant chromosome numbers. *New Phytol.* **2015**, *206*, 19–26. [[CrossRef](#)] [[PubMed](#)]
25. Greilhuber, J.; Tensch, E.M.; Loureiro, J.C.M. Nuclear DNA Content Measurement. In *Flow Cytometry with Plant Cells*; Doležel, J., Greilhuber, J., Suda, J., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2007; pp. 67–101. ISBN 9783527314874.
26. Doležel, J.; Bartoš, J. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* **2005**, *95*, 99–110. [[CrossRef](#)] [[PubMed](#)]
27. Leitch, I.J.; Bennett, M.D. Genome size and its uses: The impact of flow cytometry. In *Flow Cytometry with Plant Cells: Analysis of Genes, Chromosomes and Genomes*; Doležel, J., Greilhuber, J., Suda, J., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2007; pp. 153–176.
28. Galbraith, D.W.; Harkins, K.R.; Maddox, J.M.; Ayres, N.M.; Sharma, D.P.; Firoozabady, E. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **1983**, *220*, 1049–1051. [[CrossRef](#)] [[PubMed](#)]
29. Smith, T.W.; Kron, P.; Martin, S.L. flowPloidy: An R package for genome size and ploidy assessment of flow cytometry data. *Appl. Plant Sci.* **2018**, *6*, e01164. [[CrossRef](#)] [[PubMed](#)]
30. Doležel, J.; Bartoš, J.; Voglmayr, H.; Greilhuber, J. Nuclear DNA content and genome size of trout and human. *Cytometry* **2003**, *51A*, 127–128. [[CrossRef](#)] [[PubMed](#)]
31. Barow, M.; Meister, A. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant Cell Environ.* **2003**, *26*, 571–584. [[CrossRef](#)]
32. Barow, M.; Jovtchev, G. Endopolyploidy in Plants and its Analysis by Flow Cytometry. In *Flow Cytometry with Plant Cells*; Doležel, J., Greilhuber, J., Suda, J., Eds.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2007; pp. 349–372. ISBN 9783527314874.
33. Doležel, J.; Kubaláková, M.; Suchánková, P.; Kovářová, P.; Bartoš, J.; Šimková, H. Chromosome analysis and sorting. In *Flow Cytometry with Plant Cells*; Doležel, J., Greilhuber, J., Suda, J., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, Germany, 2007; pp. 373–404. ISBN 9783527314874.

34. Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2016**, *44*, D67–D72. [[CrossRef](#)]
35. Heap, I. The International Survey of Herbicide Resistant Weeds. Available online: www.weedscience.org (accessed on 4 January 2018).
36. United State Department of Agriculture Federal Noxious Weeds. Available online: <https://plants.usda.gov/java/noxious> (accessed on 25 July 2019).
37. Australian Government Weeds of National Significance. Available online: <https://www.environment.gov.au/biodiversity/invasive/weeds/weeds/lists/wons.html> (accessed on 25 July 2019).
38. Weber, E.; Gut, D. A survey of weeds that are increasingly spreading in Europe. In *Agronomy for Sustainable Development*; Springer Verlag/EDP Sciences/INRA: Berlin/Heidelberg, Germany, 2005; pp. 109–121.
39. Minister of Agriculture and Agri-Food Canada (AAFC). *Weed Seeds Order*; Minister of Agriculture and Agri-Food (AAFC), Canada: Ottawa, ON, Canada, 2016.
40. Straub, S.C.K.; Cronn, R.C.; Edwards, C.; Fishbein, M.; Liston, A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (*Apocynaceae*). *Genome Biol. Evol.* **2013**, *5*, 1872–1885. [[CrossRef](#)]
41. Byrne, S.L.; Erthmann, P.Ø.; Agerbirk, N.; Bak, S.; Hauser, T.P.; Nagy, I.; Paina, C.; Asp, T. The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. *Sci. Rep.* **2017**, *7*, 1–14. [[CrossRef](#)] [[PubMed](#)]
42. Bettgenhaeuser, J.; Corke, F.M.K.; Opanowicz, M.; Green, P.; Hernández-Pinzón, I.; Doonan, J.H.; Moscou, M.J. Natural Variation in *Brachypodium* Links Vernalization and Flowering Time Loci as Major Flowering Determinants. *Plant Physiol.* **2017**, *173*, 256–268. [[CrossRef](#)] [[PubMed](#)]
43. Cai, C.; Wang, X.; Liu, B.; Wu, J.; Liang, J.; Cui, Y.; Cheng, F.; Wang, X. *Brassica rapa* Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. *Mol. Plant* **2017**, *10*, 649–651. [[CrossRef](#)] [[PubMed](#)]
44. Van Bakel, H.; Stout, J.M.; Cote, A.G.; Tallon, C.M.; Sharpe, A.G.; Hughes, T.R.; Page, J.E. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* **2011**, *12*, R102. [[CrossRef](#)] [[PubMed](#)]
45. Kasianov, A.S.; Klepikova, A.V.; Kulakovskiy, I.V.; Gerasimov, E.S.; Fedotova, A.V.; Besedina, E.G.; Kondrashov, A.S.; Logacheva, M.D.; Penin, A.A. High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. *Plant J.* **2017**, *91*, 278–291. [[CrossRef](#)] [[PubMed](#)]
46. Ye, G.; Zhang, H.; Chen, B.; Nie, S.; Liu, H.; Gao, W.; Wang, H.; Gao, Y.; Gu, L. *De novo* genome assembly of the stress tolerant forest species *Casuarina equisetifolia* provides insight into secondary growth. *Plant J.* **2019**, *97*, 779–794. [[CrossRef](#)]
47. Griesmann, M.; Chang, Y.; Liu, X.; Song, Y.; Haberer, G.; Crook, M.B.; Billault-Penneteau, B.; Lauressergues, D.; Keller, J.; Imanishi, L.; et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **2018**, *361*, eaat1743. [[CrossRef](#)]
48. Wang, L.; He, F.; Huang, Y.; He, J.; Yang, S.; Zeng, J.; Deng, C.; Jiang, X.; Fang, Y.; Wen, S.; et al. Genome of Wild Mandarin and Domestication History of Mandarin. *Mol. Plant* **2018**, *11*, 1024–1037. [[CrossRef](#)]
49. Peng, Y.; Lai, Z.; Lane, T.; Nageswara-Rao, M.; Okada, M.; Jasieniuk, M.; O’Geen, H.; Kim, R.W.; Sammons, R.D.; Rieseberg, L.H.; et al. *De novo* genome assembly of the economically important weed horseweed using integrated data from multiple sequencing platforms. *Plant Physiol.* **2014**, *166*, 1241–1254. [[CrossRef](#)]
50. Sarkar, D.; Mahato, A.K.; Satya, P.; Kundu, A.; Singh, S.; Jayaswal, P.K.; Singh, A.; Bahadur, K.; Pattnaik, S.; Singh, N.; et al. The draft genome of *Corchorus olitorius* cv. JRO-524 (Navin). *Genom. Data* **2017**, *12*, 151–154. [[CrossRef](#)]
51. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Mir, G.; Gonzalez, V.M.; Henaff, E.; Camara, F.; Cozzuto, L.; Lowy, E.; et al. The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11872–11877. [[CrossRef](#)] [[PubMed](#)]
52. Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; et al. The genome sequence of the outbreeding globe artichoke constructed *de novo* incorporating a phase-aware low-pass sequencing strategy of F1 progeny. *Sci. Rep.* **2016**, *6*, 1–17.

53. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.; Iovene, M.; Sanseverino, W.; Cavagnaro, P.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657–666. [[CrossRef](#)] [[PubMed](#)]
54. Guo, L.; Qiu, J.; Ye, C.; Jin, G.; Mao, L.; Zhang, H.; Yang, X.; Peng, Q.; Wang, Y.; Jia, L.; et al. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat. Commun.* **2017**, *8*, 1–10. [[CrossRef](#)] [[PubMed](#)]
55. Badouin, H.; Gouzy, J.; Grassa, C.J.; Murat, F.; Staton, S.E.; Cottret, L.; Lelandais-Brière, C.; Owens, G.L.; Carrère, S.; Mayjonade, B.; et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **2017**, *546*, 148–152. [[CrossRef](#)] [[PubMed](#)]
56. Wu, S.; Lau, K.H.; Cao, Q.; Hamilton, J.P.; Sun, H.; Zhou, C.; Eserman, L.; Gemenet, D.C.; Olukolu, B.A.; Wang, H.; et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat. Commun.* **2018**, *9*, 1–12. [[CrossRef](#)] [[PubMed](#)]
57. Honig, J.A.; Zelzion, E.; Wagner, N.E.; Kubik, C.; Averello, V.; Vaiciunas, J.; Bhattacharya, D.; Bonos, S.A.; Meyer, W.A. Microsatellite identification in perennial ryegrass using next-generation sequencing. *Crop Sci.* **2017**, *57*, S-331–S-340. [[CrossRef](#)]
58. Vining, K.J.; Johnson, S.R.; Ahkami, A.; Lange, I.; Parrish, A.N.; Trapp, S.C.; Croteau, R.B.; Straub, S.C.K.; Pandelova, I.; Lange, B.M. Draft Genome Sequence of *Mentha longifolia* and Development of Resources for Mint Cultivar Improvement. *Mol. Plant* **2017**, *10*, 323–339. [[CrossRef](#)]
59. Zou, C.; Li, L.; Miki, D.; Li, D.; Tang, Q.; Xiao, L.; Rajput, S.; Deng, P.; Peng, L.; Jia, W.; et al. The genome of broomcorn millet. *Nat. Commun.* **2019**, *10*, 491–500. [[CrossRef](#)]
60. Guo, L.; Guo, L.; Winzer, T.; Yang, X.; Li, Y.; Ning, Z.; He, Z.; Teodor, R.; Lu, Y.; Tim, A.; et al. The opium poppy genome and morphinan production. *Science* **2018**, *362*, 343–347. [[CrossRef](#)]
61. Liu, Y.-J.; Wang, X.-R.; Zeng, Q.-Y. *De novo* assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Sci. China Life Sci.* **2019**, *62*, 609–618. [[CrossRef](#)] [[PubMed](#)]
62. Moghe, G.D.; Hufnagel, D.E.; Tang, H.B.; Xiao, Y.L.; Dworkin, I.; Town, C.D.; Conner, J.K.; Shiu, S.H. Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish *Raphanus raphanistrum* and Three Other Brassicaceae Species. *Plant Cell* **2014**, *26*, 1925–1937. [[CrossRef](#)] [[PubMed](#)]
63. Xiaohui, Z.; Zhen, Y.; Shiyong, M.; Yang, Q.; Xinhua, Y.; Xiaohua, C.; Feng, C.; Zhangyan, W.; Yuyan, S.; Yi, J.; et al. A *de novo* Genome of a Chinese Radish Cultivar. *Hortic. Plant J.* **2015**, *1*, 155–164.
64. Nakamura, N.; Hirakawa, H.; Sato, S.; Otagaki, S.; Matsumoto, S.; Tabata, S.; Tanaka, Y. Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* **2018**, *25*, 113–121. [[CrossRef](#)] [[PubMed](#)]
65. Zhang, J.; Zhang, X.; Tang, H.; Zhang, Q.; Hua, X.; Ma, X.; Zhu, F.; Jones, T.; Zhu, X.; Bowers, J.; et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **2018**, *50*, 1565–1573. [[CrossRef](#)] [[PubMed](#)]
66. Bauer, E.; Schmutzer, T.; Barilar, I.; Mascher, M.; Gundlach, H.; Martis, M.M.; Twardziok, S.O.; Hackauf, B.; Gordillo, A.; Wilde, P.; et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J.* **2017**, *89*, 853–869. [[CrossRef](#)] [[PubMed](#)]
67. Giolai, M.; Paajanen, P.; Verweij, W.; Witek, K.; Jones, J.D.G.; Clark, M.D. Comparative analysis of targeted long read sequencing approaches for characterization of a plant’s immune receptor repertoire. *BMC Genom.* **2017**, *18*, 1–15. [[CrossRef](#)] [[PubMed](#)]
68. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556. [[CrossRef](#)]
69. Dorn, K.M.; Fankhauser, J.D.; Wyse, D.L.; Marks, M.D. A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res.* **2015**, *22*, 121–131. [[CrossRef](#)]
70. Creber, H.M.C.; Davies, M.S.; Francis, D.; Walker, H.D. Variation in DNA C value in natural populations of *Dactylis glomerata* L. *New Phytol.* **1994**, *128*, 555–561. [[CrossRef](#)]
71. Beck, J. Meiotic Chromosome Counting in Flowering Plants Part 1 [Video File]. Available online: www.youtube.com/watch?v=iXqni6knH5A&t (accessed on 30 May 2019).
72. Beck, J. Meiotic Chromosome Counting in Flowering Plants Part 2 [Video File]. Available online: www.youtube.com/watch?v=xVV4qBfSQLs&t (accessed on 30 May 2019).

73. Kato, A. Air drying method using nitrous oxide for chromosome counting in maize. *Biotech. Histochem.* **1999**, *74*, 160–166. [[CrossRef](#)] [[PubMed](#)]
74. Kato, A.; Lamb, J.C.; Albert, P.S.; Danilova, T.; Han, F.; Gao, Z.; Findley, S.; Birchler, J.A. Chromosome Painting for Plant Biotechnology. In *Plant Chromosome Engineering. Methods in Molecular Biology (Methods and Protocols)*; Birchler, J.A., Ed.; Humana Press: Totowa, NJ, USA, 2011; Volume 701, pp. 67–96. ISBN 9781617379574.
75. Mandáková, T.; Lysak, M.A. Chromosome Preparation for Cytogenetic Analyses in *Arabidopsis*. *Curr. Protoc. Plant Biol.* **2016**, *1*, 43–51.
76. Chikhi, R.; Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **2014**, *30*, 31–37. [[CrossRef](#)] [[PubMed](#)]
77. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)] [[PubMed](#)]
78. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 378–390. [[CrossRef](#)] [[PubMed](#)]
79. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; MacAs, J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **2013**, *29*, 792–793. [[CrossRef](#)] [[PubMed](#)]
80. Martin, S.L.; Smith, T.; James, T.; Shalabi, F.; Kron, P.; Sauder, C.A. An update to the Canadian range and abundance of *Camelina* spp. (Brassicaceae) east of the Rocky Mountains. *Botany* **2017**, *95*, 405–417. [[CrossRef](#)]
81. Roessler, K.; Muyle, A.; Diez, C.M.; Gaut, G.R.J.; Bousios, A.; Stitzer, M.C.; Seymour, D.K.; Doebley, J.F.; Liu, Q.; Gaut, B.S. The Genomics of Selfing in Maize (*Zea mays* ssp. *mays*): Catching Purgings in the Act. *bioRxiv* **2019**, 594812. [[CrossRef](#)]
82. Palmer, C.E.D.; Keller, W.A. Overview of Haploidy. In *Biotechnology in Agriculture and Forestry Haploids in Crop Improvement II Vol.56*; Palmer, C.E., Keller, W.A., Kasha, K.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; Volume 56, pp. 3–9.
83. Forster, B.P.; Thomas, W.T.B. Doubled haploids in genetics and plant breeding. In *Plant Breeding Reviews Vol. 25*; Janick, J., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005; Volume 25, pp. 57–88. ISBN 9780471666936.
84. Dunwell, J.M. Haploids in flowering plants: Origins and exploitation. *Plant Biotechnol. J.* **2010**, *8*, 377–424. [[CrossRef](#)] [[PubMed](#)]
85. Kyriakidou, M.; Tai, H.H.; Anglin, N.L.; Ellis, D.; Strömviik, M.V. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* **2018**, *9*, 1660–1675. [[CrossRef](#)]
86. Carter, R.; Bryson, C.T.; Darbyshire, S.J. Preparation and Use of Voucher Specimens for Documenting Research in Weed Science. *Weed Technol.* **2007**, *21*, 1101–1108. [[CrossRef](#)]
87. Hussing, C.; Kampmann, M.L.; Mogensen, H.S.; Børsting, C.; Morling, N. Comparison of techniques for quantification of next-generation sequencing libraries. *Forensic Sci. Int. Genet. Suppl. Ser.* **2015**, *5*, e276–e278. [[CrossRef](#)]
88. Kreiner, J.M.; Giacomini, D.; Bemm, F.; Waithaka, B.; Regalado, J.; Lanz, C.; Hildebrandt, J.; Sikkema, P.H.; Tranel, P.J.; Weigel, D.; et al. Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *bioRxiv* **2018**, 1–17. [[CrossRef](#)]
89. Patterson, E.L.; Saski, C.A.; Sloan, D.B.; Tranel, P.J.; Westra, P.; Gaines, T.A. The draft genome of *Kochia scoparia* and the mechanism of glyphosate resistance via transposon-mediated EPSPS tandem gene duplication. *bioRxiv* **2019**. [[CrossRef](#)] [[PubMed](#)]
90. Doyle, J.J.; Doyle, J.L. A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
91. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21–29. [[CrossRef](#)] [[PubMed](#)]
92. Lander, E.S.; Waterman, M.S. Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* **1988**, *2*, 231–239. [[CrossRef](#)]
93. Soorni, A.; Haak, D.; Zaitlin, D.; Bombarely, A. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genom.* **2017**, *18*, 49–57. [[CrossRef](#)]
94. De Coster, W.; D’Hert, S.; Schultz, D.T.; Cruets, M.; Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669. [[CrossRef](#)]

95. Salmela, L.; Rivals, E. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* **2014**, *30*, 3506–3514. [[CrossRef](#)]
96. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
97. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **2012**, *1*, 18–24. [[CrossRef](#)] [[PubMed](#)]
98. Jackman, S.D.; Vandervalk, B.P.; Mohamadi, H.; Chu, J.; Yeo, S.; Hammond, S.A.; Jahesh, G.; Khan, H.; Coombe, L.; Warren, R.L.; et al. ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **2017**, *27*, 768–777. [[CrossRef](#)] [[PubMed](#)]
99. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *bioRxiv* **2016**, *27*, 722–736. [[CrossRef](#)] [[PubMed](#)]
100. Chin, C.S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [[CrossRef](#)] [[PubMed](#)]
101. Laetsch, D.R.; Blaxter, M.L. BlobTools: Interrogation of genome assemblies. *F1000Research* **2017**, *6*, 1287. [[CrossRef](#)]
102. Kajitani, R.; Yoshimura, D.; Okuno, M.; Minakuchi, Y.; Kagoshima, H.; Fujiyama, A.; Kubokawa, K.; Kohara, Y.; Toyoda, A.; Itoh, T. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat. Commun.* **2019**, *10*, 1–15. [[CrossRef](#)]
103. Waterhouse, R.M.; Seppey, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **2017**, *35*, 543–548. [[CrossRef](#)]
104. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
105. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
106. Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T.L.; Agarwala, R.; Schäffer, A.A. Database indexing for production MegaBLAST searches. *Bioinformatics* **2008**, *24*, 1757–1764. [[CrossRef](#)] [[PubMed](#)]
107. Alonge, M.; Soyk, S.; Ramakrishnan, S.; Wang, X.; Goodwin, S.; Sedlazeck, F.J.; Lippman, Z.B.; Schatz, M.C. Fast and accurate reference-guided scaffolding of draft genomes. *bioRxiv* **2019**, 519637. [[CrossRef](#)]
108. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)] [[PubMed](#)]
109. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genome Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [[CrossRef](#)]
110. Vaser, R.; Nagarajan, N.; Sović, I.; Šikic, M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 1–10. [[CrossRef](#)] [[PubMed](#)]
111. Oddes, S.; Zelig, A.; Kaplan, N. Three invariant Hi-C interaction patterns: Applications to genome assembly. *bioRxiv* **2018**, *142*, 89–99. [[CrossRef](#)] [[PubMed](#)]
112. Ghurye, J.; Pop, M.; Koren, S.; Bickhart, D.; Chin, C.S. Scaffolding of long read assemblies using long range contact information. *BMC Genom.* **2017**, *18*, 527–538. [[CrossRef](#)] [[PubMed](#)]
113. Kronenberg, Z.N.; Rhie, A.; Koren, S.; Concepcion, G.T.; Peluso, P.; Munson, K.M.; Hiendleder, S.; Fedrigo, O.; Jarvis, E.D.; Adam, M.; et al. Extended haplotype phasing of de novo genome assemblies with FALCON-Phase. *bioRxiv* **2018**, 1–27. [[CrossRef](#)]
114. Jibrán, R.; Dzierzon, H.; Bassil, N.; Bushakra, J.M.; Edger, P.P.; Sullivan, S.; Finn, C.E.; Dossett, M.; Vining, K.J.; Vanburen, R.; et al. Chromosome-scale scaffolding of the black raspberry (*Rubus occidentalis* L.) genome based on chromatin interaction data. *Hortic. Res.* **2018**, *5*, 8–19. [[CrossRef](#)]
115. Lightfoot, D.J.; Jarvis, D.E.; Ramaraj, T.; Lee, R.; Jellen, E.N.; Maughan, P.J. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.* **2017**, *15*, 74. [[CrossRef](#)]

116. Appels, R.; Eversole, K.; Feuillet, C.; Keller, B.; Rogers, J.; Stein, N.; Ronen, G. International Wheat Genome Sequencing Consortium Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[PubMed](#)]
117. Deschamps, S.; Zhang, Y.; Llaca, V.; Ye, L.; Sanyal, A.; King, M.; May, G.; Lin, H. A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *Nat. Commun.* **2018**, *9*, 4844. [[CrossRef](#)] [[PubMed](#)]
118. Edger, P.P.; Poorten, T.J.; VanBuren, R.; Hardigan, M.A.; Colle, M.; McKain, M.R.; Smith, R.D.; Teresi, S.J.; Nelson, A.D.L.; Wai, C.M.; et al. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **2019**, *51*, 541–547. [[CrossRef](#)] [[PubMed](#)]
119. Bosi, E.; Donati, B.; Galardini, M.; Brunetti, S.; Sagot, M.-F.; Lió, P.; Crescenzi, P.; Fani, R.; Fondi, M. MeDuSa: A multi-draft based scaffold. *Bioinformatics* **2015**, *31*, 2443–2451. [[CrossRef](#)] [[PubMed](#)]
120. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [[CrossRef](#)] [[PubMed](#)]
121. Stanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004**, *32*, W309–W312. [[CrossRef](#)] [[PubMed](#)]
122. Volfovsky, N.; Hass, B.J.; Salzberg, S.L. A clustering method for repeat analysis in DNA sequences. *Genome Biol.* **2001**, *2*, 0027.1–0027.11. [[CrossRef](#)] [[PubMed](#)]
123. Shi, J.; Liang, C. Generic Repeat Finder: A high-sensitivity tool for genome-wide de novo repeat detection. *Plant Physiol.* **2019**, *180*, 1803–1815. [[CrossRef](#)] [[PubMed](#)]
124. Mao, H.; Wang, H. SINE-scan: An efficient tool to discover short interspersed nuclear elements (SINEs) in large-scale genomic datasets. *Bioinformatics* **2017**, *33*, 743–745. [[PubMed](#)]
125. Chen, J.; Hu, Q.; Zhang, Y.; Lu, C.; Kuang, H. P-MITE: A database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res.* **2014**, *42*, 1176–1181. [[CrossRef](#)] [[PubMed](#)]
126. Xiong, W.; He, L.; Lai, J.; Dooner, H.K.; Du, C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10263–10268. [[CrossRef](#)] [[PubMed](#)]
127. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [[CrossRef](#)] [[PubMed](#)]
128. Campbell, M.S.; Law, M.; Holt, C.; Stein, J.C.; Moghe, G.D.; Hufnagel, D.E.; Lei, J.; Achawanantakun, R.; Jiao, D.; Lawrence, C.J.; et al. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiol.* **2014**, *164*, 513–524. [[CrossRef](#)] [[PubMed](#)]
129. Audano, P.A.; Sulovari, A.; Graves-Lindsay, T.A.; Cantilieri, S.; Sorensen, M.; Welch, A.M.E.; Dougherty, M.L.; Nelson, B.J.; Shah, A.; Dutcher, S.K.; et al. Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* **2019**, *176*, 663–675. [[CrossRef](#)] [[PubMed](#)]
130. Hackl, T.; Hedrich, R.; Schultz, J.; Förster, F. Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **2014**, *30*, 3004–3011. [[CrossRef](#)] [[PubMed](#)]
131. Gnerre, S.; Maccallum, I.; Przybylski, D.; Ribeiro, F.J.; Burton, J.N.; Walker, B.J.; Sharpe, T.; Hall, G.; Shea, T.P.; Sykes, S.; et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1513–1518. [[CrossRef](#)] [[PubMed](#)]
132. English, A.C.; Richards, S.; Han, Y.; Wang, M.; Vee, V.; Qu, J.; Qin, X.; Muzny, D.M.; Reid, J.G.; Worley, K.C.; et al. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE* **2012**, *7*, e47768. [[CrossRef](#)] [[PubMed](#)]
133. Mayela Soto-Jimenez, L.; Estrada, K.; Sanchez-Flores, A. GARM: Genome assembly, reconciliation and merging. *Curr. Top. Med. Chem.* **2014**, *14*, 418–424. [[CrossRef](#)] [[PubMed](#)]
134. Linthorst, J.; Hulsman, M.; Holstege, H.; Reinders, M. Scalable multi whole-genome alignment using recursive exact matching. *bioRxiv* **2015**, 022715. [[CrossRef](#)]
135. Gao, S.; Bertrand, D.; Chia, B.K.H.; Nagarajan, N. OPERA-LG: Efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol.* **2016**, *17*, 102. [[CrossRef](#)]
136. Lukashin, A.V.; Borodovsky, M. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **1998**, *26*, 1107–1115. [[CrossRef](#)] [[PubMed](#)]
137. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **2006**, *7*, S10. [[CrossRef](#)] [[PubMed](#)]

138. Petit, S.; Boursault, A.; Le Guilloux, M.; Munier-Jolain, N.; Reboud, X. Weeds in agricultural landscapes. A review. *Agron. Sustain. Dev.* **2011**, *31*, 309–317. [[CrossRef](#)]
139. De Wet, J.M.J.; Harlan, J.R. Weeds and Domesticates: Evolution in the man-made habitat. *Econ. Bot.* **1975**, *29*, 99–107. [[CrossRef](#)]
140. Barrett, S.H. Crop mimicry in weeds. *Econ. Bot.* **1983**, *37*, 255–282. [[CrossRef](#)]
141. Harlan, J.R.; de Wet, J.M.J. Some thoughts about weeds. *Econ. Bot.* **1965**, *19*, 16–24. [[CrossRef](#)]
142. Tedin, O. Vererbung, Variation Und Systeme-Matik in Der Gattung Camelina (German with English Summary). *Hereditas* **1925**, *6*, 275–386. [[CrossRef](#)]
143. Powles, S.B.; Yu, Q. Evolution in action: Plants resistant to herbicides. *Annu. Rev. Plant Biol.* **2010**, *61*, 317–347. [[CrossRef](#)]
144. Neve, P.; Vila-Aiub, M.; Roux, F. Evolutionary-thinking in agricultural weed management. *New Phytol.* **2009**, *184*, 783–793. [[CrossRef](#)]
145. Baker, H.G. The evolution of weeds. *Annu. Rev. Ecol. Syst.* **1974**, *5*, 1–24. [[CrossRef](#)]
146. Warwick, S.I.; Stewart, C.N., Jr. Crops come from wild plants: How domestication, transgenes, and linkage together shape ferality. In *Crop Ferality and Volunteerism*; CRC Press: Boca Raton, FL, USA, 2005; pp. 9–30. ISBN 9783540773405.
147. Ellstrand, N.C.; Heredia, S.M.; Leak-Garcia, J.A.; Heraty, J.M.; Burger, J.C.; Yao, L.; Nohzadeh-Malakshah, S.; Ridley, C.E. Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evol. Appl.* **2010**, *3*, 494–504. [[CrossRef](#)] [[PubMed](#)]
148. Vigueira, C.C.; Olsen, K.M.; Caicedo, A.L. The red queen in the corn: Agricultural weeds as models of rapid adaptive evolution. *Heredity* **2013**, *110*, 303–311. [[CrossRef](#)] [[PubMed](#)]
149. Royal, C.D.; Novembre, J.; Fullerton, S.M.; Goldstein, D.B.; Long, J.C.; Bamshad, M.J.; Clark, A.G. Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *Am. J. Hum. Genet.* **2010**, *86*, 661–673. [[CrossRef](#)] [[PubMed](#)]
150. Kidd, K.K.; Speed, W.C.; Pakstis, A.J.; Furtado, M.R.; Fang, R.; Madbouly, A.; Maiers, M.; Middha, M.; Friedlaender, F.R.; Kidd, J.R. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci. Int.* **2014**, *10*, 23–32. [[CrossRef](#)] [[PubMed](#)]
151. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [[PubMed](#)]
152. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application. *Genetics* **1992**, *131*, 479–491. [[PubMed](#)]
153. Pickrell, J.K.; Pritchard, J.K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **2012**, *8*, e1002967. [[CrossRef](#)] [[PubMed](#)]
154. Slatkin, M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **1993**, *47*, 264–279. [[CrossRef](#)] [[PubMed](#)]
155. Vekemans, X.; Hardy, O.J. New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol. Ecol.* **2004**, *13*, 921–935. [[CrossRef](#)] [[PubMed](#)]
156. Aguillon, S.M.; Fitzpatrick, J.W.; Bowman, R.; Schoech, S.J.; Clark, A.G.; Coop, G.; Chen, N. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. *PLoS Genet.* **2017**, *13*, e1006911. [[CrossRef](#)] [[PubMed](#)]
157. Bradburd, G.S.; Coop, G.M.; Ralph, P.L. Inferring continuous and discrete population genetic structure across space. *Genetics* **2018**, *210*, 33–52. [[CrossRef](#)] [[PubMed](#)]
158. Lawson, D.J.; van Dorp, L.; Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **2018**, *9*, 3258. [[CrossRef](#)] [[PubMed](#)]
159. Gutenkunst, R.N.; Hernandez, R.D.; Williamson, S.H.; Bustamante, C.D. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet.* **2009**, *5*, e1000695. [[CrossRef](#)] [[PubMed](#)]
160. Excoffier, L.; Dupanloup, I.; Huerta-Sánchez, E.; Sousa, V.C.; Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.* **2013**, *9*, e1003905. [[CrossRef](#)] [[PubMed](#)]
161. Smith, J.M.; Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res. Camb.* **1974**, *23*, 23–35. [[CrossRef](#)]
162. Przeworski, M. The Signature of Positive Selection at Randomly Chosen Loci. *Genetics* **2002**, *160*, 1179–1189. [[PubMed](#)]

163. Fay, J.C.; Wu, C.-I. Hitchhiking Under Positive Darwinian Selection. *Genetics* **2000**, *155*, 1405–1413. [[PubMed](#)]
164. Hermisson, J.; Pennings, P.S. Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation. *Genetics* **2005**, *169*, 2335–2352. [[CrossRef](#)]
165. Meirmans, P.G.; Hedrick, P.W. Assessing population structure: F_{ST} and related measures. *Mol. Ecol. Resour.* **2011**, *11*, 5–18. [[CrossRef](#)]
166. Wright, S. Genetical Structure of Populations. *Nature* **1950**, 247–249. [[CrossRef](#)]
167. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595. [[PubMed](#)]
168. Fay, J.C.; Wu, C.-I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* **2003**, *4*, 213–235. [[CrossRef](#)] [[PubMed](#)]
169. Pavlidis, P.; Živković, D.; Stamatakis, A.; Alachiotis, N. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol. Biol. Evol.* **2013**, *30*, 2224–2234. [[CrossRef](#)] [[PubMed](#)]
170. Degiorgio, M.; Huber, C.D.; Hubisz, M.J.; Hellmann, I.; Nielsen, R. SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **2016**, *32*, 1895–1897. [[CrossRef](#)]
171. Nielsen, R.; Williamson, S.; Kim, Y.; Hubisz, M.J.; Clark, A.G.; Bustamante, C. Genomic scans for selective sweeps using SNP data. *Genome Res.* **2005**, *15*, 1566–1575. [[CrossRef](#)] [[PubMed](#)]
172. Berg, J.J.; Coop, G. A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.* **2014**, *10*, e1004412. [[CrossRef](#)] [[PubMed](#)]
173. Berg, J.J.; Harpak, A.; Sinnott-armstrong, N.; Joergensen, A.M.; Mostafavi, H.; Field, Y.; Boyle, E.A.; Zhang, X.; Racimo, F.; Pritchard, J.K.; et al. Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* **2019**, *8*, e39725. [[CrossRef](#)] [[PubMed](#)]
174. Mosyakin, S.L.; Robertson, K.R. *Amaranthus*. In *Flora of North America North of Mexico*; Flora of North America Editorial Committee: New York, NY, USA; Oxford, UK, 2004.
175. Délye, C.; Menchari, Y.; Michel, S.; Cadet, E.; Le Corre, V. A new insight into arable weed adaptive evolution: Mutations endowing herbicide resistance also affect germination dynamics and seedling emergence. *Ann. Bot.* **2013**, *111*, 681–691. [[CrossRef](#)]
176. Baucom, R.S. Evolutionary and ecological insights from herbicide-resistant weeds: What have we learned about plant adaptation, and what is left to uncover? *New Phytol.* **2019**, *223*, 68–82. [[CrossRef](#)]
177. Yuan, J.S.; Tranel, P.J.; Stewart, C.N. Non-target-site herbicide resistance: A family business. *Trends Plant Sci.* **2007**, *12*, 6–13. [[CrossRef](#)]
178. Délye, C.; Jasieniuk, M.; Le Corre, V. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **2013**, *29*, 649–658. [[CrossRef](#)] [[PubMed](#)]
179. Beckie, H.J.; Tardif, F.J. Herbicide cross resistance in weeds. *Crop Prot.* **2012**, *35*, 15–28. [[CrossRef](#)]
180. Délye, C. Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: A major challenge for weed science in the forthcoming decade. *Pest Manag. Sci.* **2013**, *62*, 176–187. [[CrossRef](#)] [[PubMed](#)]
181. Ghanizadeh, H.; Harrington, K.C. Non-target Site Mechanisms of Resistance to Herbicides. *CRC. Crit. Rev. Plant Sci.* **2017**, *36*, 24–34. [[CrossRef](#)]
182. Yang, Q.; Liu, Y.J.; Zeng, Q.Y. Overexpression of three orthologous glutathione S-transferases from *Populus* increased salt and drought resistance in *Arabidopsis*. *Biochem. Syst. Ecol.* **2019**, *83*, 57–61. [[CrossRef](#)]
183. Conte, S.S.; Lloyd, A.M. Exploring multiple drug and herbicide resistance in plants—Spotlight on transporter proteins. *Plant Sci.* **2011**, *180*, 196–203. [[CrossRef](#)] [[PubMed](#)]
184. Cummins, I.; Wortley, D.J.; Sabbadin, F.; He, Z.; Coxon, C.R.; Straker, H.E.; Sellars, J.D.; Knight, K.; Edwards, L.; Hughes, D.; et al. Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5812–5817. [[CrossRef](#)] [[PubMed](#)]
185. Van Etten, M.; Lee, K.M.; Chang, S.-M.; Baucom, R.S. Parallel and nonparallel genomic responses contribute to herbicide resistance in *Ipomoea purpurea*, a common agricultural weed. *bioRxiv* **2019**. [[CrossRef](#)]
186. Salas-Perez, R.A.; Sasaki, C.A.; Noorai, R.E.; Srivastava, S.K.; Lawton-Rauh, A.L.; Nichols, R.L.; Roma-Burgos, N. RNA-Seq transcriptome analysis of *Amaranthus palmeri* with differential tolerance to glufosinate herbicide. *PLoS ONE* **2018**, *13*, e0195488. [[CrossRef](#)]
187. Bai, S.; Liu, W.; Wang, H.; Zhao, N.; Jia, S.; Zou, N.; Guo, W.; Wang, J. Enhanced herbicide metabolism and metabolic resistance genes identified in tribenuron-methyl resistant *Myosoton aquaticum* L. *J. Agric. Food Chem.* **2018**, *66*, 9850–9857. [[CrossRef](#)]

188. Kreuz, K.; Tommasini, R.; Martinoia, E. Old Enzymes for a New Job Herbicide Detoxification in Plants. *Plant Physiol.* **1996**, *111*, 349–353. [[CrossRef](#)] [[PubMed](#)]
189. Edwards, R.; Del Buono, D.; Fordham, M.; Skipsey, M.; Brazier, M.; Dixon, D.P.; Cummins, I. Differential induction of glutathione transferases and glucosyltransferases in wheat, maize and *Arabidopsis thaliana* by herbicide safeners. *Z. Nat.* **2005**, *60*, 307–316. [[CrossRef](#)] [[PubMed](#)]
190. Schuler, M.A.; Werck-Reichhart, D. Functional Genomics of P450s. *Annu. Rev. Plant Biol.* **2003**, *54*, 629–667. [[CrossRef](#)] [[PubMed](#)]
191. Nelson, D.R.; Schuler, M.A.; Paquette, S.M.; Werck-Reichhart, D.; Bak, S. Comparative Genomics of Rice and *Arabidopsis*. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot. *Plant Physiol.* **2004**, *135*, 756–772. [[CrossRef](#)] [[PubMed](#)]
192. Ehlting, J.; Provart, N.J.; Werck-Reichhart, D. Functional annotation of the *Arabidopsis* P450 superfamily based on large-scale co-expression analysis. *Biochem. Soc. Trans.* **2006**, *34*, 1192–1198. [[CrossRef](#)] [[PubMed](#)]
193. Werck-Reichhart, D.; Hehn, A.; Diderjean, L. Cytochromes P450 for engineering herbicide tolerance. *Plant Cell* **2004**, *5*, 116–123. [[CrossRef](#)]
194. Inui, H.; Ueyama, Y.; Shiota, N.; Ohkawa, Y.; Ohkawa, H. Herbicide Metabolism and Cross-Tolerance in Transgenic Potato Plants Expressing Human CYP1A1. *Pestic. Biochem. Physiol.* **1999**, *64*, 33–46. [[CrossRef](#)]
195. Kawahigashi, H.; Hirose, S.; Ohkawa, H.; Ohkawa, Y. Herbicide resistance of transgenic rice plants expressing human CYP1A1. *Biotechnol. Adv.* **2007**, *25*, 75–84. [[CrossRef](#)] [[PubMed](#)]
196. Kawahigashi, H.; Hirose, S.; Ohkawa, H.; Ohkawa, Y. Phytoremediation of the herbicides atrazine and metolachlor by transgenic rice plants expressing human CYP1A1, CYP2B6, and CYP2C19. *J. Agric. Food Chem.* **2006**, *54*, 2985–2991. [[CrossRef](#)]
197. Hirose, S.; Kawahigashi, H.; Ozawa, K.; Shiota, N.; Inui, H.; Ohkawa, H.; Ohkawa, Y. Transgenic rice containing human CYP2B6 detoxifies various classes of herbicides. *J. Agric. Food Chem.* **2005**, *53*, 3461–3467. [[CrossRef](#)]
198. Herbicide Resistance Action Committee. Available online: <https://www.hracglobal.com/> (accessed on 26 January 2019).
199. Robineau, T.; Batard, Y.; Nedelkina, S.; Cabello-Hurtado, F.; LeRet, M.; Sorokine, O.; Didierjean, L.; Werck-Reichhart, D. The Chemically Inducible Plant Cytochrome P450 CYP76B1 Actively Metabolizes Phenylureas and Other Xenobiotics. *Plant Physiol.* **2002**, *118*, 1049–1056. [[CrossRef](#)] [[PubMed](#)]
200. Siminszky, B.; Corbin, F.T.; Ward, E.R.; Fleischmann, T.J.; Dewey, R.E. Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1750–1755. [[CrossRef](#)] [[PubMed](#)]
201. Höfer, R.; Boachon, B.; Renault, H.; Gavira, C.; Miesch, L.; Iglesias, J.; Ginglinger, J.-F.; Allouche, L.; Miesch, M.; Grec, S.; et al. Dual Function of the Cytochrome P450 CYP76 Family from *Arabidopsis thaliana* in the Metabolism of Monoterpenols and Phenylurea Herbicides. *Plant Physiol.* **2014**, *166*, 1149–1161. [[CrossRef](#)] [[PubMed](#)]
202. Khanom, S.; Jang, J.; Lee, O.R. Overexpression of ginseng cytochrome P450 CYP736A12 alters plant growth and confers phenylurea herbicide tolerance in *Arabidopsis*. *J. Ginseng Res.* **2019**. [[CrossRef](#)]
203. Iwakami, S.; Endo, M.; Saika, H.; Okuno, J.; Nakamura, N.; Yokoyama, M.; Watanabe, H.; Toki, S.; Uchino, A.; Inamura, T. Cytochrome P450 CYP81A12 and CYP81A21 Are Associated with Resistance to Two Acetolactate Synthase Inhibitors in *Echinochloa phyllopogon*. *Plant Physiol.* **2014**, *165*, 618–629. [[CrossRef](#)] [[PubMed](#)]
204. Guo, F.; Iwakami, S.; Yamaguchi, T.; Uchino, A.; Sunohara, Y.; Matsumoto, H. Role of CYP81A cytochrome P450s in clomazone metabolism in *Echinochloa phyllopogon*. *Plant Sci.* **2019**, *283*, 321–328. [[CrossRef](#)] [[PubMed](#)]
205. Yang, Q.; Li, J.; Shen, J.; Xu, Y.; Liu, H.; Deng, W.; Li, X.; Zheng, M. Metabolic Resistance to Acetolactate Synthase Inhibiting Herbicide Tribenuron-Methyl in *Descurainia sophia* L. Mediated by Cytochrome P450 Enzymes. *J. Agric. Food Chem.* **2018**, *66*, 4319–4327. [[CrossRef](#)] [[PubMed](#)]
206. Oliveira, M.C.; Gaines, T.A.; Dayan, F.E.; Patterson, E.L.; Jhala, A.J.; Knezevic, S.Z. Reversing resistance to tembotrione in an *Amaranthus tuberculatus* (var. *rudis*) population from Nebraska, USA with cytochrome P450 inhibitors. *Pest Manag. Sci.* **2018**, *74*, 2296–2305. [[CrossRef](#)]
207. Hidayat, I.; Preston, C. Cross-resistance to imazethapyr in a fluazifop-P-butyl-resistant population of *Digitaria sanguinalis*. *Pestic. Biochem. Physiol.* **2001**, *71*, 190–195. [[CrossRef](#)]
208. Marrs, K.A. The Functions and Regulation of Glutathione S-Transferases in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 127–158. [[CrossRef](#)]

209. Stavridou, E.; Voulgari, G.; Bosmali, I.; Chronopoulou, E.G.; Lo Cicero, L.; Lo Piero, A.R.; Labrou, N.E.; Tsaftaris, A.; Nianiou-Obeidat, I.; Madesis, P. Plant Adaptation to Stress Conditions: The Case of Glutathione S-Transferases (GSTs). In *Biotic and Abiotic Stress Tolerance in Plants*; Vats, S., Ed.; Springer Nature Singapore Pte Ltd.: Berlin/Heidelberg, Germany, 2018; pp. 173–202. ISBN 978-981-10-9028-8.
210. Labrou, N.E.; Papageorgiou, A.C.; Pavli, O.; Flemetakis, E. Plant GSTome: Structure and functional role in xenome network and plant stress response. *Curr. Opin. Biotechnol.* **2015**, *32*, 186–194. [[CrossRef](#)] [[PubMed](#)]
211. Li, L.; Hou, M.; Cao, L.; Xia, Y.; Shen, Z.; Hu, Z. Glutathione S-transferases modulate Cu tolerance in *Oryza sativa*. *Environ. Exp. Bot.* **2018**, *155*, 313–320. [[CrossRef](#)]
212. Li, Z.-K.; Chen, B.; Li, X.-X.; Wang, J.-P.; Zhang, Y.; Wang, X.-F.; Yan, Y.-Y.; Ke, H.F.; Yang, J.; Wu, J.-H.; et al. A newly identified cluster of glutathione S-transferase genes provides *Verticillium* wilt resistance in cotton. *Plant J.* **2019**, *98*, 213–227. [[CrossRef](#)] [[PubMed](#)]
213. Dixon, D.P.; Laphorn, A.; Edwards, R. Protein family review Plant glutathione transferases. *Genome Biol.* **2002**, *3*, 1–10. [[CrossRef](#)] [[PubMed](#)]
214. Dixon, D.; Cole, D.J.; Edwards, R. Characterisation of multiple glutathione transferases containing the GST I subunit with activities toward herbicide substrates in maize (*Zea mays*). *Pestic. Sci.* **1997**, *50*, 72–82. [[CrossRef](#)]
215. Grove, G.; Zarlengo, R.P.; Timmerman, K.P.; Li, N.Q.; Tam, M.F.; Tu, C.-P.D. Characterization and heterospecific expression of cDNA clones of genes in the maize GSH S-transferase multigene family. *Nucleic Acids Res.* **1988**, *16*, 425–438. [[CrossRef](#)] [[PubMed](#)]
216. Karavangeli, M.; Labrou, N.E.; Clonis, Y.D.; Tsaftaris, A. Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation. *Biomol. Eng.* **2005**, *22*, 121–128. [[CrossRef](#)]
217. Milligan, A.S.; Daly, A.; Parry, M.A.J.; Lazzeri, P.A.; Jepson, I. The expression of a maize glutathione S-transferase gene in transgenic wheat confers herbicide tolerance, both in planta and in vitro. *Mol. Breed.* **2001**, *7*, 301–315. [[CrossRef](#)]
218. Benekos, K.; Kissoudis, C.; Nianiou-Obeidat, I.; Labrou, N.; Madesis, P.; Kalamaki, M.; Makris, A.; Tsaftaris, A. Overexpression of a specific soybean GmGSTU4 isoenzyme improves diphenyl ether and chloroacetanilide herbicide tolerance of transgenic tobacco plants. *J. Biotechnol.* **2010**, *150*, 195–201. [[CrossRef](#)]
219. Cummins, I.; Cole, D.J.; Edwards, R. A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *Plant J.* **1999**, *18*, 285–292. [[CrossRef](#)]
220. Petit, C.; Duhieu, B.; Boucansaud, K.; Délye, C. Complex genetic control of non-target-site-based resistance to herbicides inhibiting acetyl-coenzyme A carboxylase and acetolactate-synthase in *Alopecurus myosuroides* Huds. *Plant Sci.* **2010**, *178*, 501–509. [[CrossRef](#)]
221. Wright, A.A.; Rodriguez-Carres, M.; Sasidharan, R.; Koski, L.; Peterson, D.G.; Nandula, V.K.; Ray, J.D.; Bond, J.A.; Shaw, D.R. Multiple Herbicide-Resistant Junglerice (*Echinochloa colona*): Identification of Genes Potentially Involved in Resistance through Differential Gene Expression Analysis. *Weed Sci.* **2018**, *66*, 347–354. [[CrossRef](#)]
222. Nakka, S.; Godar, A.S.; Thompson, C.R.; Peterson, D.E.; Jugulam, M. Rapid detoxification via glutathione S-transferase (GST) conjugation confers a high level of atrazine resistance in Palmer amaranth (*Amaranthus palmeri*). *Pest Manag. Sci.* **2017**, *73*, 2236–2243. [[CrossRef](#)] [[PubMed](#)]
223. Dücker, R.; Zöllner, P.; Lümmen, P.; Ries, S.; Collavo, A.; Beffa, R. Glutathione transferase plays a major role in flufenacet resistance of ryegrass (*Lolium* spp.) field populations. *Pest Manag. Sci.* **2019**. [[CrossRef](#)]
224. Balabanova, D.; Remans, T.; Vassilev, A.; Cuypers, A.; Vangronsveld, J. Possible involvement of glutathione S-transferases in imazamox detoxification in an imidazolinone-resistant sunflower hybrid. *J. Plant Physiol.* **2018**, *221*, 62–65. [[CrossRef](#)] [[PubMed](#)]
225. Sharma, R.; Draicchio, F.; Bull, H.; Herzig, P.; Maurer, A.; Pillen, K.; Thomas, W.T.B.; Flavell, A.J. Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. *J. Exp. Bot.* **2018**, *69*, 3811–3822. [[CrossRef](#)] [[PubMed](#)]
226. Krajewski, M.P.; Kanawati, B.; Fekete, A.; Kowalski, N.; Schmitt-Kopplin, P.; Grill, E. Analysis of *Arabidopsis* glutathione-transferases in yeast. *Phytochemistry* **2013**, *91*, 198–207. [[CrossRef](#)] [[PubMed](#)]
227. Theodoulou, F.L. Plant ABC transporters. *Biochim. Biophys. Acta Biomembr.* **2000**, *1465*, 79–103. [[CrossRef](#)]
228. Hwang, J.U.; Song, W.Y.; Hong, D.; Ko, D.; Yamaoka, Y.; Jang, S.; Yim, S.; Lee, E.; Khare, D.; Kim, K.; et al. Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle. *Mol. Plant* **2016**, *9*, 338–355. [[CrossRef](#)] [[PubMed](#)]

229. Sánchez-Fernández, R.; Davies, T.G.E.; Coleman, J.O.D.; Rea, P.A. The *Arabidopsis thaliana* ABC Protein Superfamily, a Complete Inventory. *J. Biol. Chem.* **2001**, *276*, 30231–30244. [[CrossRef](#)] [[PubMed](#)]
230. Ofori, P.A.; Mizuno, A.; Suzuki, M.; Martinoia, E.; Reuscher, S.; Aoki, K.; Shibata, D.; Otagaki, S.; Matsumoto, S.; Shiratake, K. Genome-wide analysis of atp binding cassette (ABC) transporters in tomato. *PLoS ONE* **2018**, *13*, e0200854. [[CrossRef](#)] [[PubMed](#)]
231. Busi, R.; Goggin, D.E.; Heap, I.M.; Horak, M.J.; Jugulam, M.; Masters, R.A.; Napier, R.M.; Riar, D.S.; Satchivi, N.M.; Torra, J.; et al. Weed resistance to synthetic auxin herbicides. *Pest Manag. Sci.* **2018**, *74*, 2265–2276. [[CrossRef](#)] [[PubMed](#)]
232. Peng, Y.; Abercrombie, L.L.G.; Yuan, J.S.; Riggins, C.W.; Sammons, R.D.; Tranel, P.J.; Stewart, C.N., Jr.; Characterization of the horseweed (*Conyza canadensis*) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes. *Pest Manag. Sci.* **2010**, *66*, 1053–1062. [[CrossRef](#)] [[PubMed](#)]
233. Windsor, B.; Roux, S.J.; Lloyd, A. Multiherbicide tolerance conferred by AtPgp1 and apyrase overexpression in *Arabidopsis thaliana*. *Nat. Biotechnol.* **2003**, *21*, 428–433. [[CrossRef](#)] [[PubMed](#)]
234. Jo, J.; Won, S.H.; Son, D.; Lee, B.H. Paraquat resistance of transgenic tobacco plants over-expressing the *Ochrobactrum anthropi pqrA* gene. *Biotechnol. Lett.* **2004**, *26*, 1391–1396. [[CrossRef](#)] [[PubMed](#)]
235. Quistgaard, E.M.; Löw, C.; Guettou, F.; Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 123–132. [[CrossRef](#)]
236. Ward, J.M. Identification of novel families of membrane proteins from the model plant *Arabidopsis thaliana*. *Bioinformatics* **2001**, *17*, 560–563. [[CrossRef](#)]
237. Teixeira, M.C.; Duque, P.; Sá-Correia, I. Environmental genomics: Mechanistic insights into toxicity of and resistance to the herbicide 2,4-D. *Trends Biotechnol.* **2007**, *25*, 363–370. [[CrossRef](#)]
238. Cabrito, T.R.; Teixeira, M.C.; Duarte, A.A.; Duque, P.; Sá-Correia, I. Heterologous expression of a Tpo1 homolog from *Arabidopsis thaliana* confers resistance to the herbicide 2,4-D and other chemical stresses in yeast. *Appl. Microbiol. Biotechnol.* **2009**, *84*, 927–936. [[CrossRef](#)]
239. Tiwari, P.; Sangwan, R.S.; Sangwan, N.S. Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnol. Adv.* **2016**, *34*, 714–739. [[CrossRef](#)]
240. Pflugmacher, S.; Schröder, P.; Sandermann Jr, H. Taxonomic distribution of plant glutathione S-transferases acting on xenobiotics. *Phytochemistry* **2000**, *54*, 267–273. [[CrossRef](#)]
241. Brazier-Hicks, M.; Offen, W.A.; Gershater, M.C.; Revett, T.J.; Lim, E.-K.; Bowles, D.J.; Davies, G.J.; Edwards, R. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20238–20243. [[CrossRef](#)] [[PubMed](#)]
242. Loutre, C.; Dixon, D.P.; Brazier, M.; Slater, M.; Cole, D.J.; Edwards, R. Isolation of a glucosyltransferase from *Arabidopsis thaliana* active in the metabolism of the persistent pollutant 3,4-dichloroaniline. *Plant J.* **2003**, *34*, 485–493. [[CrossRef](#)] [[PubMed](#)]
243. Wetzel, A.; Sandermann, H. Plant biochemistry of xenobiotics: Isolation and characterization of a soybean O-glucosyltransferase of DDT metabolism. *Arch. Biochem. Biophys.* **1994**, *314*, 323–328. [[CrossRef](#)] [[PubMed](#)]
244. Brazier-Hicks, M.; Edwards, R. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant J.* **2005**, *42*, 556–566. [[CrossRef](#)] [[PubMed](#)]
245. Meßner, B.; Thulke, O.; Schäffner, A.R. *Arabidopsis* glucosyltransferases with activities toward both endogenous and xenobiotic substrates. *Planta* **2003**, *217*, 138–146.
246. Cha, J.-Y.; Lee, D.-Y.; Ali, I.; Jeong, S.Y.; Shin, B.; Ji, H.; Kim, J.S.; Kim, M.-G.; Kim, W.-Y. *Arabidopsis* GIGANTEA negatively regulates chloroplast biogenesis and resistance to herbicide butafenacil. *Plant Cell Rep.* **2019**, *38*, 793–801. [[CrossRef](#)]
247. Etter, P.D.; Bassham, S.; Hohenlohe, P.A.; Johnson, E.; Cresko, W.A. SNP Discovery and genotyping for evolutionary genetics using RAD Sequencing. In *Molecular Methods for Evolutionary Genetics*; Orgogozo, V., Rockman, M.V., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2011; Volume 772, pp. 1–9. ISBN 978-1-61779-227-4.
248. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135. [[CrossRef](#)]
249. Catchen, J.M.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [[CrossRef](#)]

250. Rochette, N.C.; Catchen, J.M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **2017**, *12*, 2640–2659. [CrossRef]
251. Paris, J.R.; Stevens, J.R.; Catchen, J.M. Lost in parameter space: A road map for stacks. *Methods Ecol. Evol.* **2017**, *8*, 1360–1373. [CrossRef]
252. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
253. R Core Team, R. *A Language and Environment for Statistical Computing*; R Core Team R: Vienna, Austria, 2017.
254. Paradis, E.; Gosselin, T.; Goudet, J.; Jombart, T.; Schliep, K. Linking genomics and population genetics with R. *Mol. Ecol. Resour.* **2017**, *17*, 54–66. [CrossRef] [PubMed]
255. Goudet, J.; Jombart, T. *Hierfstat: Estimation and Tests of Hierarchical F-Statistics*; CRAN: 2015; p. 58. Available online: <https://cran.r-project.org/web/packages/hierfstat/index.html> (accessed on 30 May 2019).
256. Pembleton, L.W.; Cogan, N.O.I.; Forster, J.W. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* **2013**, *13*, 946–952. [CrossRef] [PubMed]
257. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction v2.6.1. *PeerJ* **2018**, *2*, e281. [CrossRef]
258. Revell, L.J. phytools: Phylogenetic tools for comparative biology (and other things). *Methods Ecol. Evol.* **2012**, *3*, 217–223. [CrossRef]
259. Jombart, T.; Ahmed, I. ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 1403. [CrossRef] [PubMed]
260. Küpper, A.; Manmathan, H.K.; Giacomini, D.; Patterson, E.L.; McCloskey, W.B.; Gaines, T.A. Population Genetic Structure in Glyphosate-Resistant and -Susceptible Palmer Amaranth (*Amaranthus palmeri*) Populations Using Genotyping-by-sequencing (GBS). *Front. Plant Sci.* **2018**, *9*, 29. [CrossRef]
261. Foll, M.; Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **2008**, *180*, 977–993. [CrossRef]
262. Günther, T.; Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **2013**, *195*, 205–220. [CrossRef]
263. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.M.; Birol, I. ABySS: A parallel assembler for short read sequence data. *Genome Res.* **2009**, *19*, 1117–1123. [CrossRef] [PubMed]
264. Ye, C.; Hill, C.M.; Wu, S.; Ruan, J.; Ma, Z. (Sam) DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* **2016**, *6*, 1–9.
265. Hoshino, A.; Jayakumar, V.; Nitasaka, E.; Toyoda, A.; Noguchi, H.; Itoh, T.; Shin, T.; Minakuchi, Y.; Koda, Y.; Nagano, A.J.; et al. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat. Commun.* **2016**, *7*, 1–10. [CrossRef] [PubMed]
266. Pickar-Oliver, A.; Gersbach, C.A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **2019**. [CrossRef] [PubMed]
267. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–822. [CrossRef] [PubMed]
268. Doudna, J.A.; Charpentier, E. The new frontier of genome engineering with CRISPR–Cas9. *Science* **2014**, *346*, 1258096. [CrossRef] [PubMed]
269. Neve, P. Gene drive systems: Do they have a place in agricultural weed management? *Pest Manag. Sci.* **2018**, *74*, 2671–2679. [CrossRef] [PubMed]
270. Bull, J.J.; Malik, H.S. The gene drive bubble: New realities. *PLOS Genet.* **2017**, *13*, e1006850. [CrossRef] [PubMed]
271. Committee on Gene Drive Research in Non-Human Organisms: Recommendations for Responsible Conduct; Board on Life Sciences; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine. *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*; The National Academies Press: Washington, DC, USA, 2016; ISBN 9780309437875.
272. Courtier-Orgogozo, V.; Morizot, B.; Boëte, C. Agricultural pest control with CRISPR-based gene drive: Time for public debate. *EMBO Rep.* **2017**, *18*, 878–880. [CrossRef]
273. Esvelt, K.M.; Smidler, A.L.; Catteruccia, F.; Church, G.M. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **2014**, e03401. [CrossRef]

274. Champer, J.; Buchman, A.; Akbari, O.S. Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **2016**, *17*, 146–159. [[CrossRef](#)] [[PubMed](#)]
275. Knoll, A.; Fauser, F.; Puchta, H. DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosom. Res.* **2014**, *22*, 191–201. [[CrossRef](#)] [[PubMed](#)]
276. Huang, T.K.; Puchta, H. CRISPR/Cas-mediated gene targeting in plants: Finally a turn for the better for homologous recombination. *Plant Cell Rep.* **2019**, *38*, 443–453. [[CrossRef](#)] [[PubMed](#)]
277. Klompe, S.E.; Vo, P.L.H.; Halpin-Healy, T.S.; Sternberg, S.H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **2019**, *571*, 219–225. [[CrossRef](#)] [[PubMed](#)]
278. Que, Q.; Chilton, M.D.M.; Elumalai, S.; Zhong, H.; Dong, S.; Shi, L. Repurposing macromolecule delivery tools for plant genetic modification in the era of precision genome engineering. *Methods Mol. Biol.* **2019**, *1864*, 3–18. [[PubMed](#)]
279. Grunwald, H.A.; Gantz, V.M.; Poplawski, G.; Xu, X.-R.S.; Bier, E.; Cooper, K.L. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **2019**, *566*, 105–109. [[CrossRef](#)]
280. Charpentier, M.; Khedher, A.H.; Menoret, S.; Brion, A.; Lamribet, K.; Dardillac, E.; Boix, C.; Perrouault, L.; Tesson, L.; Geny, S.; et al. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* **2018**, *9*, 1–11. [[CrossRef](#)]
281. Wang, M.; Lu, Y.; Botella, J.R.; Mao, Y.; Hua, K.; Zhu, J. Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Mol. Plant* **2017**, *10*, 1007–1010. [[CrossRef](#)]
282. Dahan-Meir, T.; Filler-Hayut, S.; Melamed-Bessudo, C.; Bocobza, S.; Czosnek, H.; Aharoni, A.; Levy, A.A. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *Plant J.* **2018**, *95*, 5–16. [[CrossRef](#)]
283. Gil-Humanes, J.; Wang, Y.; Liang, Z.; Shan, Q.; Ozuna, C.V.; Sánchez-León, S.; Baltés, N.J.; Starker, C.; Barro, F.; Gao, C.; et al. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* **2017**, *89*, 1251–1262. [[CrossRef](#)]
284. Kyrou, K.; Hammond, A.M.; Galizi, R.; Kranjc, N.; Burt, A.; Beaghton, A.K.; Nolan, T.; Crisanti, A. A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* **2018**, *36*, 1062–1066. [[CrossRef](#)] [[PubMed](#)]
285. Webber, B.L.; Raghu, S.; Edwards, O.R. Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10565–10567. [[CrossRef](#)] [[PubMed](#)]
286. Baltzegar, J.; Cavin Barnes, J.; Elsensohn, J.E.; Gutzmann, N.; Jones, M.S.; King, S.; Sudweeks, J. Anticipating complexity in the deployment of gene drive insects in agriculture. *J. Responsible Innov.* **2018**, *5*, S81–S97. [[CrossRef](#)]
287. Oye, K.A.; Esvelt, K.; Appleton, E.; Catteruccia, F.; Church, G.; Kuiken, T.; Lightfoot, S.B.-Y.; McNamara, J.; Smidler, A.; Collins, J.P. Regulating gene drives. *Science* **2014**, *345*, 626–628. [[CrossRef](#)] [[PubMed](#)]
288. Unckless, R.L.; Messer, P.W.; Connallon, T.; Clark, A.G. Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics* **2015**, *201*, 425–431. [[CrossRef](#)] [[PubMed](#)]
289. Unckless, R.L.; Clark, A.G.; Messer, P.W. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **2017**, *205*, 827–841. [[CrossRef](#)] [[PubMed](#)]
290. Champer, J.; Reeves, R.; Oh, S.Y.; Liu, C.; Liu, J.; Clark, A.G.; Messer, P.W. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **2017**, *13*, e1006796. [[CrossRef](#)] [[PubMed](#)]
291. Noble, C.; Olejarz, J.; Esvelt, K.M.; Church, G.M.; Nowak, M.A. Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **2017**, *3*, e1601964. [[CrossRef](#)] [[PubMed](#)]
292. Hemingway, J.; Hawkes, N.J.; McCarroll, L.; Ranson, H. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem. Mol. Biol.* **2004**, *34*, 653–665. [[CrossRef](#)] [[PubMed](#)]
293. Ranson, H.; Lissenden, N. Insecticide Resistance in African *Anopheles* Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends Parasitol.* **2016**, *32*, 187–196. [[CrossRef](#)] [[PubMed](#)]
294. Cohuet, A.; Harris, C.; Robert, V.; Fontenille, D. Evolutionary forces on *Anopheles*: What makes a malaria vector? *Trends Parasitol.* **2010**, *26*, 130–136. [[CrossRef](#)] [[PubMed](#)]
295. Hemingway, J.; Ranson, H.; Magill, A.; Kolaczinski, J.; Fornadel, C.; Gimnig, J.; Coetzee, M.; Simard, F.; Roch, D.K.; Hinzoumbe, C.K.; et al. Averting a malaria disaster: Will insecticide resistance derail malaria control? *Lancet* **2016**, *387*, 1785–1788. [[CrossRef](#)]

296. Namountougou, M.; Simard, F.; Baldet, T.; Diabaté, A.; Ouédraogo, J.B.; Martin, T.; Dabiré, R.K. Multiple Insecticide Resistance in *Anopheles gambiae* s.l. Populations from Burkina Faso, West Africa. *PLoS ONE* **2012**, *7*, e48412. [[CrossRef](#)] [[PubMed](#)]
297. Edi, C.V.A.; Koudou, B.G.; Jones, C.M.; Weetman, D.; Ranson, H. Multiple-Insecticide Resistance in *Anopheles gambiae* Mosquitoes, Southern Côte d'Ivoire. *Emerg. Infect. Dis.* **2012**, *18*, 1508–1511. [[CrossRef](#)] [[PubMed](#)]
298. Hammond, A.; Galizi, R.; Kyrou, K.; Simoni, A.; Siniscalchi, C.; Katsanos, D.; Gribble, M.; Baker, D.; Marois, E.; Russell, S.; et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **2016**, *34*, 78–83. [[CrossRef](#)]
299. Holt, R.A.; Broder, S.; Subramanian, G.M.; Halpern, A.L.; Sutton, G.G.; Charlab, R.; Nusskern, D.R.; Wincker, P.; Clark, A.G.; Ribeiro, J.M.C.; et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **2002**, *298*, 129–149. [[CrossRef](#)] [[PubMed](#)]
300. Scudellari, M. Hijacking evolution. *Nature* **2019**, *571*, 160–162. [[CrossRef](#)]
301. Collins, C.M.; Bonds, J.A.S.; Quinlan, M.M.; Mumford, J.D. Effects of the removal or reduction in density of the malaria mosquito, *Anopheles gambiae* s.l., on interacting predators and competitors in local ecosystems. *Med. Vet. Entomol.* **2019**, *33*, 1–15. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).