horticulturae

MDPI

Review

Integrating Genomic Markers and Non-Invasive Phenotyping for Early Sex Identification in Horticultural Plants: A Mechanism-Guided Framework

Junzhu Zou [1], Ke Shi [2], Haidong Wu [3], Hao Shen [4], Yuxiao Qu [1], Ao Li [1,*] and Junxiang Liu [1,*]

[1] Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China; zoujz@caf.ac.cn (J.Z.); quyuxiao@caf.ac.cn (Y.Q.)
[2] Beijing Jindu Landscape and Afforesting Co., Ltd., Beijing 100140, China; shike202606@163.com
[3] Information Center of Ministry of Ecology and Environment, Beijing 100029, China; wuhaidong2026@163.com
[4] School of Grassland Science, Beijing Forestry University, Beijing 100083, China; shenhao@bjfu.edu.cn
* Correspondence: liao@caf.ac.cn (A.L.); jxliu@caf.ac.cn (J.L.); Tel.: +86-10-62889652 (J.L.)

Highlights

What are the main findings?

- Sex identification technologies should be selected according to sex determination mechanisms and diagnostic targets.
- Genomic markers and non-invasive phenotyping capture complementary layers of plant sex expression.

What are the implications of the main findings?

- Mechanism-guided screening can support propagation, seedling selection, and germplasm management.
- Spectral prescreening followed by molecular confirmation may improve throughput and reliability.

Abstract

Early sex identification is essential for the propagation, cultivation, quality improvement, and germplasm management of dioecious horticultural plants and related functionally dioecious systems, particularly in perennial species with long juvenile phases. However, the reliability and transferability of sex-identification technologies depend strongly on the underlying sex-determining mechanism. Here, we synthesize recent advances in plant sex determination and diagnostic technologies, ranging from morphological and biochemical traits to molecular markers, high-throughput sequencing, structural-variant detection, and emerging non-invasive phenotyping. We propose that sex-identification strategies should be selected according to the biological target generated by each mechanism, including heteromorphic sex chromosomes, homomorphic sex-determining regions (SDRs), functional sex-determining genes, sex chromosome turnover, dosage-dependent systems, and environmentally labile sex expression. We further distinguish genetic, developmental, physiological, and phenotypic layers of plant sex, emphasizing that DNA markers and spectral phenotyping provide complementary information. Genomic markers and non-invasive phenotyping are expected to be consistent when genetic sex is stably expressed, but they may become inconsistent when sex expression is developmentally, hormonally, or environmentally modulated. While molecular markers remain the most reliable tools for confirmatory genotyping, Raman spectroscopy, surface-enhanced Raman scattering (SERS), hyperspectral imaging, and machine learning may serve as rapid prescreening tools in large breeding populations, although their application remains at the

Academic Editor: Sergey V. Dolgov

Received: 5 June 2026
Revised: 12 July 2026
Accepted: 15 July 2026
Published: 17 July 2026

proof-of-concept stage. Finally, we present a mechanism-guided decision framework for integrating genomic markers and non-invasive phenotyping to support early sex screening, propagation planning, planting-material optimization, and marker-assisted improvement in dioecious horticultural plants.

Keywords: dioecy; sex identification; sex determination region; molecular markers; non-invasive phenotyping; seedling selection

1. Introduction

Sexual system diversity in angiosperms has important implications for evolutionary biology, crop breeding, and germplasm management [1]. Although approximately 90% of angiosperm species are hermaphroditic or monoecious, roughly 6% have transitioned into distinct dioecious or functionally dioecious systems [2–4]. In these taxa, sexual dimorphism often governs economic value, agronomic management, and industrial utility across various agricultural and forestry production systems [5]. For instance, sex strongly affects production targets in fruit crops such as kiwifruit (*Actinidia* spp.) and ginkgo (*Ginkgo biloba* L.), as well as in vegetables like asparagus (*Asparagus officinalis* L.) and spinach (*Spinacia oleracea* L.), where male and female individuals may differ in fruit or seed production, edible organ quality, yield stability, or cultivation value. Conversely, in urban forestry and landscape management, male individuals of the Salicaceae family are often preferred to reduce airborne dispersal of seed hairs from females. However, dioecious woody plants typically exhibit a prolonged juvenile phase, during which sexes cannot be morphologically distinguished prior to flowering. This diagnostic delay leads to inefficient use of land, labor, and cultivation resources, thereby slowing breeding progress and reducing production efficiency. Consequently, reliable, low-cost, and early-stage sex identification technologies are increasingly important for improving breeding efficiency, cultivation planning, and germplasm utilization in dioecious plants.

Historically, early-stage sex screening relied primarily on vegetative morphological traits and physiological or biochemical markers. However, these phenotypic features display high developmental plasticity and are easily confounded by fluctuating environmental variables, leading to diagnostic instability. Recent genomic advancements have begun to unravel the diverse genetic foundations and evolutionary trajectories underlying sex determination in dioecious species, highlighting genetic transitions ranging from classic active Y-chromosome mechanisms to complex chromosome dosage balances [3,6]. As Henry et al. [7] summarized, dioecious plants have evolved "a hundred ways to make a sex," reflecting the remarkable diversity of sex-determining mechanisms in plants. These mechanisms range from the *de novo* origin of sex chromosomes and compact sex-determining regions (SDRs) to hormone-mediated regulation of floral sex expression. With the transition from anonymous markers such as RAPD and AFLP to high-throughput genomic mapping, researchers have increasingly localized SDRs and identified key regulatory genes, including ARR17 in Salicaceae and ACS11 in Cucurbitaceae models [8–11]. Recent transcriptomic and spatial transcriptomic studies further indicate that sex-related regulatory divergence can emerge before obvious morphological differentiation, as shown by early sex-biased expression in hemp (*Cannabis sativa* L.) and early-stage sex differentiation in *S. oleracea* [12,13].

Despite these genomic insights, a fundamental challenge persists regarding how to translate mechanistic knowledge into optimal diagnostic pipelines. Recent applied studies illustrate this translational challenge: multiplex PCR assays based on Y-specific

coding-region markers have achieved high-fidelity seedling sex identification in *C. sativa*, whereas genome-wide association study (GWAS)-assisted genomic selection has enabled high-precision early sex prediction in *Populus deltoides* W.Bartram ex Marshall breeding populations [14,15]. We argue that a mechanism-guided selection of identification targets significantly improves diagnostic accuracy, cross-species transferability, and practical applicability. Simultaneously, with the dawn of plant phenomics, non-destructive, real-time, and high-throughput optical technologies combined with machine learning (ML)—such as Raman spectroscopy—have demonstrated substantial potential for capturing in-field biochemical fingerprints [16]. Within this framework, spectral and ML approaches are viewed as emerging complementary tools, rather than replacements for validated molecular markers, as they mainly capture downstream physiological states, not the inherited genotype.

To define the scope of this critical review, we searched major scientific databases, including Web of Science, Scopus, PubMed, and Google Scholar. The literature search focused primarily on publications from the last decade, while earlier foundational studies were also included when necessary. Search terms covered three related domains: plant sex determination, including "sex chromosome evolution", "sex-determining region", "structural variation", "presence/absence variation", and "copy number variation"; molecular diagnostics and phenomics, including "sex-linked marker", "KASP", "hyperspectral imaging", and "Raman spectroscopy"; and integrative breeding technologies, including "machine learning", "multi-omics integration", and "genomic selection". Given the diversity of species, sex-determining mechanisms, diagnostic targets, and detection technologies, this review focuses on a mechanism-guided synthesis of representative advances. Building on this scope, we examine the evolution of early-stage sex identification technologies in dioecious plants and functionally dioecious systems. Here, "mechanism-guided" means that the underlying sex-determining mechanism defines the diagnostic target, which then guides the selection of appropriate detection technologies. We address two primary questions: (1) how sex identification technologies can be selected according to different sex determination mechanisms, and (2) how non-invasive phenotyping may complement DNA markers when genetic sex cannot fully predict realized developmental, physiological, or phenotypic sex. Specifically, we aim to:

1. link major sex-determining mechanisms in dioecious plants with corresponding identification targets;
2. evaluate the reliability and transferability of sex-identification markers, tracking the transition from loosely linked markers to functional and structural variants;
3. propose an integrated genotype–phenotype framework in which non-invasive phenotyping serves as a rapid prescreening layer and molecular markers provide confirmatory evidence.

2. Sex Determination Mechanisms as Identification Targets

Plant sex determination exhibits remarkable evolutionary diversity and complexity across angiosperms. Based on regulatory hierarchies, existing sex determination systems can be classified into several distinct evolutionary models. Elucidating these multi-layered genetic mechanisms and precisely mapping them to specific genomic or metabolic targets is a fundamental prerequisite for establishing efficient, early-stage sex identification technologies (Figure 1).

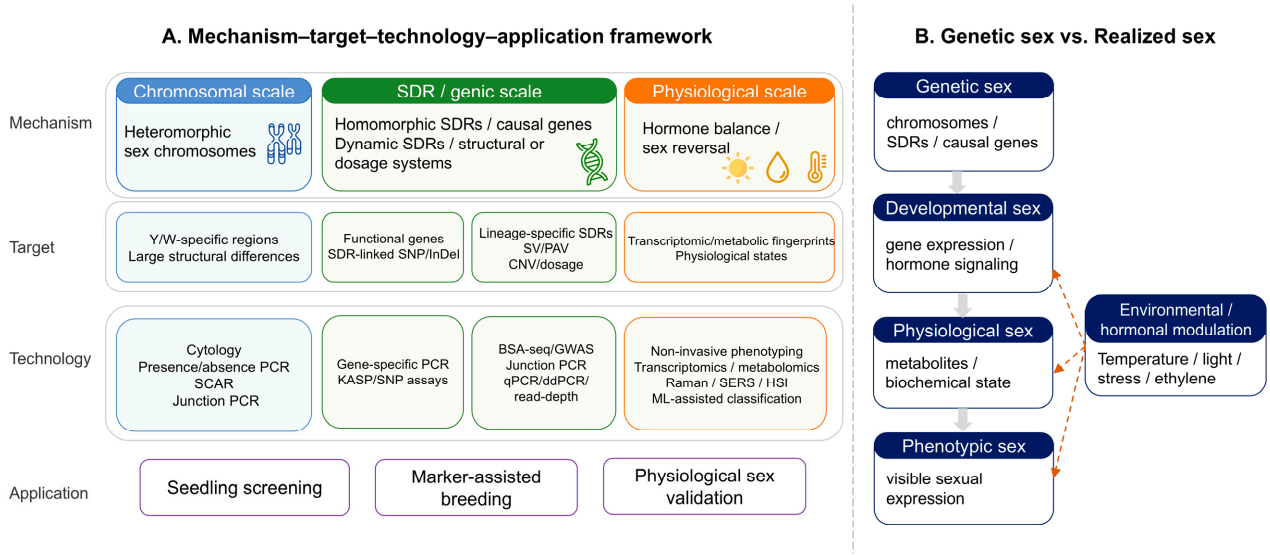

Figure 1. Conceptual framework for mechanism-guided sex identification in dioecious plants. (**A**) Sex determination mechanisms generate distinct diagnostic targets, which guide the selection of appropriate detection technologies and practical applications. Heteromorphic sex chromosomes provide chromosomal-scale structural targets; homomorphic or dynamic SDRs generate functional, lineage-specific, structural, or dosage-based targets; and environmentally labile systems may produce transcriptomic, metabolic, or physiological characteristics. (**B**) Plant sex expression can be conceptualized as a hierarchical process from genetic sex to realized developmental, physiological, and phenotypic sex. Genomic markers primarily detect inherited genetic sex, whereas non-invasive phenotyping captures downstream physiological or phenotypic sex expression. These two diagnostic layers are expected to be consistent when genetic sex is stably expressed, but may diverge when sex expression is modulated by developmental stage, hormonal regulation, environmental stress, or sex reversal, explaining why they provide complementary rather than interchangeable information. In panel (**A**), blue, green, and orange denote chromosomal-scale, SDR/genic-scale, and physiological-scale diagnostic layers, respectively, while purple denotes applications; in panel (**B**), solid gray arrows indicate the hierarchical progression from genetic sex to realized sex expression, and dashed orange arrows indicate environmental or hormonal modulation. Colors are used for conceptual grouping and do not indicate quantitative ranking.

2.1. Heteromorphic Sex Chromosomes

In dioecious plants with highly differentiated sex chromosomes, sex expression is typically associated with morphologically distinct, heteromorphic X/Y or Z/W chromosome pairs. According to classic evolutionary models, recombination suppression around a newly established sex-determining region (SDR) can maintain linkage between sex-determining loci and sexually antagonistic alleles. Over time, reduced recombination may promote the accumulation of repetitive sequences, deleterious mutations, and gene degeneration within the non-recombining Y- or W-linked region, ultimately contributing to cytological differentiation between sex chromosomes [2,3]. Classic examples are found in *Silene latifolia* Poir. and *C. sativa*, where repeat accumulation and degeneration of the non-recombining region have contributed to marked Y-chromosome enlargement relative to the X chromosome [17]. Chromosome-level assemblies have also revealed highly differentiated XY systems in marine angiosperms such as *Phyllospadix torreyi* S. Watson, suggesting that heteromorphic sex chromosomes have arisen independently across diverse plant families [18]. Diagnostically, these systems provide large sex-linked structural targets, including Y- or W-specific sequences and repeat-enriched non-recombining regions, which can be converted into presence/absence PCR assays, SCAR markers, or breakpoint-based structural assays. However, the repetitive and degenerate nature of these regions can complicate primer

design and increase the risk of non-specific amplification across diverse germplasms. These systems represent chromosomal-scale diagnostic targets in Figure 1A.

2.2. Homomorphic SDRs and Functional Sex-Determining Genes

In many dioecious angiosperms, sex chromosomes remain homomorphic or only weakly differentiated, and sex expression is governed by compact sex-determining regions (SDRs) harboring key regulatory genes. Homomorphic SDRs can involve either single-factor switches or tightly linked two-gene systems. *Actinidia* exemplifies the latter, with the Y-specific SDR containing *FrBy* to promote stamen development and *SyGI*, a Type-C response regulator, to suppress carpel development, thereby maintaining dioecy through tight genetic linkage [19]. Mechanistically, functional characterization revealed that *SyGI* arose via a lineage-specific duplication of an autosomal Type-C cytokinin response regulator, operating as a dominant suppressor of feminization [20]. Although *Actinidia* has often been treated as a stable homomorphic system, recent genomic evidence indicates recurrent neo-sex chromosome turnover in the genus, partly associated with bursts of transposable element insertions [21]. Importantly, the functional sex-determining module appears to be conserved despite changes in chromosomal context, illustrating that diagnostic targets may remain functionally stable even when SDR location shifts among lineages. Another example of a two-gene homomorphic SDR has recently been reported in *Itoa orientalis* Hemsl., where *IoTDF1* and *IoMIF2* act as male-promoting and female-suppressing factors, respectively [22].

Single-factor sex determination pathways are well illustrated in *Diospyros*, where sex determination is mediated by a Y-encoded small RNA mechanism. In diploid *Diospyros lotus* L., a Y-chromosome-specific pseudogene named *OGI* produces small RNAs that target and trans-silence the autosomal gene *MeGI*, a master regulator of floral sex fates, thereby suppressing female traits to promote maleness [23]. This genetic switch undergoes a critical evolutionary and epigenetic expansion in polyploid lineages; in hexaploid *Diospyros kaki* Thunb., alternative DNA methylation states at the *MeGI* promoter establish a flexible layer of regulation, allowing reversible monoecious or hermaphroditic phenotypes that are not determined solely by the presence of the *OGI* locus [7,24]. For these stable homomorphic systems, targeting the causal functional genes rather than flanking regions can support the development of functional or near-perfect markers in validated genetic backgrounds. High-throughput platforms such as KASP can then be used to genotype these variants for large-scale seedling screening. These examples illustrate the genic-scale pathway in Figure 1A.

2.3. Sex Chromosome Turnover and Lineage-Specific Markers

Despite the structural conservation found in some genera, homomorphic SDRs can be highly dynamic, frequently undergoing rapid sex chromosome turnovers where the master sex-determining locus translocates across different chromosomes among closely related species. This evolutionary flexibility is pronounced within the Salicaceae family. Indeed, such frequent turnovers challenge the canonical evolutionary model of expanding recombination suppression and progressive chromosome-wide degeneration; by constantly resetting the sex-determining locus to new autosomes, the sex chromosomes remain largely homomorphic and the SDR is kept compact [25]. While the sex determination locus is highly conserved on chromosome 19 in *Populus*, the ancestral SDR in *Salix* is situated on chromosome 15 [10]. Chromosome-level assemblies of *Salix dunnii* C.K.Schneid., however, revealed that its SDR has translocated to chromosome 07 [9]. Mechanistically, these turnovers are often driven by transposable elements or structural rearrangements; for instance, the newly evolved Y chromosome in *Salix exigua* Nutt. acquired an inverted

repeat of *ARR17*, which generates small interfering RNAs (siRNAs) to trans-silence the ancestral, fully functional *ARR17* locus via epistatic suppression [26]. Furthermore, in *Salix purpurea* L., sex expression involves a complex dual-gene co-regulation network mediated by both *ARR17* and *GATA15*, defying simple classification [10]. This recurrent shifting of the master switch among non-homologous chromosomes underlines why large-scale genetic decay is rarely observed in these nascent systems. A key diagnostic implication of these dynamics is that molecular markers developed for homomorphic SDRs typically suffer from poor cross-species or cross-lineage transferability. Flanking markers linked to a specific chromosome can become decoupled or obsolete due to translocation, necessitating repeated mapping of lineage-specific SDRs and the development of species-specific diagnostic toolkits. Such turnover shifts the diagnostic target itself, making species-specific SDR mapping and validation essential.

2.4. Structural Variants, PAVs and CNVs

Beyond single-nucleotide variations, recent long-read sequencing technologies have revealed that plant sex determination is frequently driven by large-scale structural variants (SVs), presence/absence variations (PAVs), and copy number variations (CNVs). In the dioecious weed *Amaranthus palmeri* S. Watson, the male-specific Y haplotype is characterized by a massive ~1.3 Mb sequence insertion PAV, the spontaneous deletion of which results in a direct phenotypic shift from dioecy to monoecy [27]. In dosage-dependent systems, sex is influenced by the relative copy number or ratio of genomic components. A representative example is *Humulus lupulus* L., where sex determination is governed by an X-chromosome-to-autosome (X:A) dosage balance rather than a single master gene [28]. Similarly, in *C. sativa*, the copy number variation of the florigen ortholog *CsFT1* not only dictates flowering time but also exhibits distinct sex-specific expression profiles [29]. SV-associated sex determination architectures have also been reported in *Silene* and tree species within Juglandaceae.

These examples highlight that sex determination can involve large-scale genomic architecture rather than only SNP- or InDel-level variation. From a diagnostic perspective, this shifts marker design toward structural or quantitative targets. Large PAVs and inversions can be interrogated using breakpoint-spanning junction PCR or related structural assays, whereas dosage-dependent systems require quantitative approaches such as qPCR, ddPCR, or read-depth-based CNV analysis. These examples extend the mechanism-guided framework to structural and dosage-based targets.

2.5. Environmentally Labile Sex Expression and Hormone-Mediated Systems

While genomic architectures define the foundational genetic sex, certain plant lineages display highly plastic, environmentally labile sex expression where the phenotypic sex can diverge from the underlying genotype under environmental or hormonal influences. At the cellular and developmental levels, plant hormone signaling networks integrate environmental cues with genetic programs and thereby influence sex differentiation [30]. *Cucumis* is included here as a model for hormone-mediated sex expression rather than as a strictly dioecious system. In Cucurbitaceae models such as *Cucumis sativus* L. and *Cucumis melo* L., the auxin response factor *CsARF3* directly binds to the promoter of the ethylene biosynthesis gene *CsACS11*, activating ethylene biosynthesis and promoting female flower development [8]. This hormone-driven sexual plasticity is not exclusively restricted to species lacking heteromorphic chromosomes. Even in species with an underlying XY system, such as *C. sativa*, multi-omic analyses suggest that hormone-responsive pathways, particularly ethylene signaling, can modulate floral sex expression and contribute to phenotypic sex reversal under certain conditions [31]. Furthermore, in *Carica papaya* L., despite

possessing a nascent, non-degenerate Y chromosome, extreme environmental fluctuations frequently trigger sex reversal; high-temperature stress suppresses carpel development, effectively transforming genetic hermaphrodites or females into phenotypic males [32].

These genotype-by-environment (G × E) interactions complicate early-stage sex screening because DNA markers primarily detect genetic sex rather than realized developmental, physiological, or phenotypic sex. In environmentally sensitive species, molecular genotyping should therefore be complemented by downstream validation tools, such as transcriptomics, metabolomics, or non-invasive spectral phenotyping. This mechanism corresponds to the layered model in Figure 1B and to the physiological-scale pathway in Figure 1A.

To synthesize the diverse sex determination mechanisms discussed above, Table 1 summarizes representative genetic systems, diagnostic targets, recommended identification strategies, and their current application readiness in dioecious or functionally dioecious plant systems. This mechanistic diversity underscores a fundamental principle of sex diagnostics: no universal marker exists across all dioecious taxa. Instead, marker design should be guided by the genetic architecture, SDR stability, and validation context of the target species.

Table 1. Representative sex determination mechanisms, diagnostic targets, and identification strategies in dioecious or functionally dioecious plant systems.

Species/Group	Sex System	Key Determinant/Target	Recommended Identification Strategy	Diagnostic Readiness	Key References
Actinidia	XY (Homomorphic)	*FrBy*/*SyGI* functional genes	Gene-specific PCR/KASP assays	High in validated *Actinidia* backgrounds	[19–21]
Diospyros	XY (Homomorphic)	*OGI*/*MeGI* small-RNA pathway	Gene-specific markers/Epigenetic validation	Moderate–high	[23,24]
Salicaceae	XY/ZW; dynamic SDR turnover	*ARR17*-related SDR, SVs, siRNA-mediated regulation	Species-specific SDR mapping and SV markers	Moderate	[25,26,33]
Humulus	X:A dosage balance	Chromosome dosage/CNV	qPCR/ddPCR/read-depth CNV	Emerging	[6,28]
Silene	XY (Heteromorphic)	Y-linked SDR/repeat-rich non-recombining region	Y-specific markers/structural assays	High for genetic sex	[3,17]
Amaranthus	XY (PAV-driven)	Male-specific insertion PAV (~1.3 Mb)	PAV-targeted PCR/Junction PCR	High for genetic sex	[27]
Cannabis	XY with environmentally/hormonally modulated sex expression	Y-linked regions; hormone- or metabolite-associated signatures	DNA markers + Raman/hyperspectral phenotyping	Moderate for DNA; emerging for spectral phenotyping	[31,34]
Cucumis/*Cucurbitaceae models*	Monoecious/gynoecious systems; hormone-mediated sex expression	Ethylene synthesis pathway (*ACS* genes)	Marker-assisted selection for flower sex expression traits	High for Cucurbitaceae breeding models	[8,30]
C. papaya	XY/XY/XY^h (Prone to environmental sex reversal)	Nascent Y/SVs + Temperature stress	SDR markers + phenotypic validation under environmental stress	Moderate–high	[32,35]

Note: Diagnostic readiness indicates the current maturity of each strategy for practical sex identification. High refers to systems in which causal genes, SDRs, or sex-linked structural variants have been relatively well characterized and can support reliable marker development in validated genetic backgrounds; Moderate refers to systems with validated markers or candidate loci but limited cross-population or cross-species transferability; Emerging refers to systems in which the underlying mechanism has been recently resolved, proposed, or only partially translated into diagnostic applications. Cucumis is included as a representative model for hormone-mediated floral sex expression rather than as a strictly dioecious system. For *C. papaya*, Y^h denotes the hermaphrodite-specific Y chromosome; XX, XY, and XY^h correspond to female, male, and hermaphrodite genotypes, respectively.

3. Molecular Marker Systems and Diagnostic Reliability

Before DNA-based assays became widely available, early sex identification relied on morphological, physiological, or biochemical traits. For example, vegetative sexual dimorphism, such as differences in petiole length and stem diameter, has been explored in *C. papaya* before the onset of reproductive development [36]. More recent phenotype-based approaches have incorporated computer vision and machine learning to extract geometric

features from seedling leaves, achieving up to 88% diagnostic accuracy in *Pistacia vera* L. [37], while electrochemical sensing platforms have also been developed for metabolic sex discrimination in *Taxus* species [38]. However, these traits are often developmentally delayed, environmentally sensitive, and difficult to transfer across populations. This limitation motivated a shift toward molecular marker systems that detect more stable genetic targets associated with sex determination. In this section, we focus on how DNA-based markers have evolved from anonymous linked fragments to SDR-linked, functional, structural, and dosage-based diagnostic assays. The historical trajectory of sex-identification technologies is summarized in Figure 2, which links early phenotype-based screening, DNA marker development, genomics-guided diagnostics, and emerging phenomics-assisted prescreening.

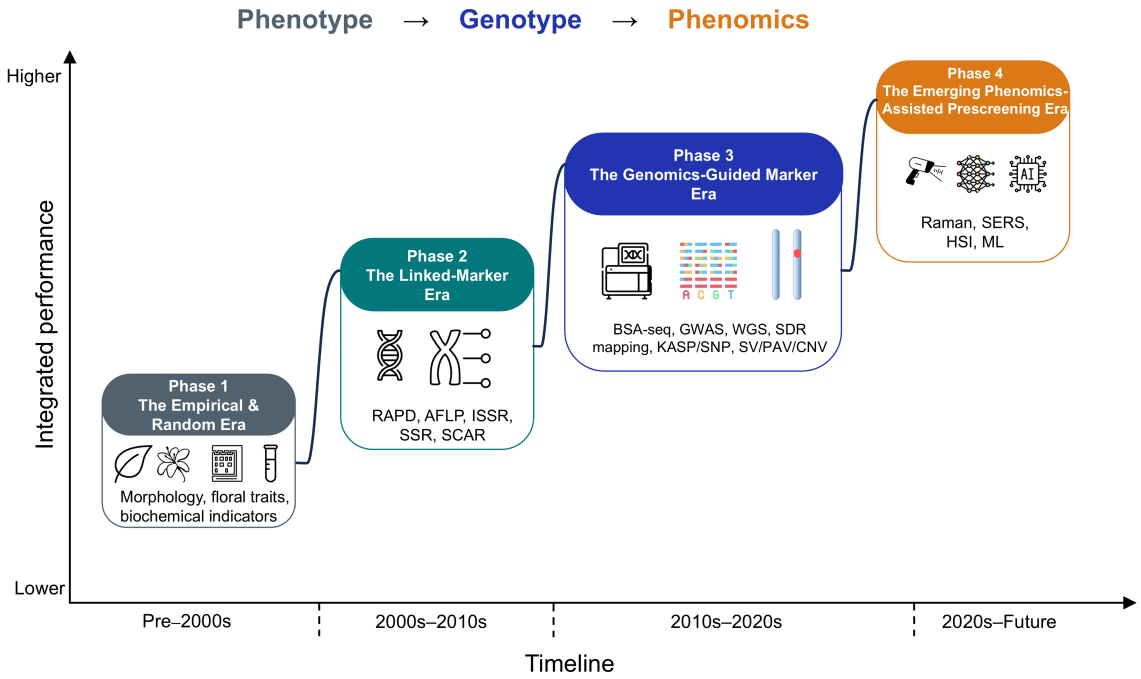

Figure 2. Evolution of sex-identification technologies in dioecious plants. Sex-identification technologies have progressed from empirical phenotype-based screening and historical linked markers to genomics-guided marker development and emerging phenomics-assisted prescreening. The trajectory highlights a shift from macroscopic phenotype observation to genetic-marker-based diagnostics and, more recently, to high-dimensional optical phenotyping. Spectral and machine-learning approaches are positioned as complementary prescreening tools rather than replacements for molecular confirmation.

3.1. Historical Markers: RAPD, AFLP, ISSR, SSR and SCAR

The first generation of molecular sex identification relied on random amplification techniques such as RAPD, AFLP, and ISSR, which required little or no prior genomic sequence information. Early workflows successfully identified sex-linked diagnostic bands in species such as *Encephalartos natalensis* R.A.Dyer & I.Verd. [39], while the integration of AFLP with bulked segregant analysis (BSA) enabled the discovery of male-specific markers in *Piper betle* L. [40]. Despite their historical utility, these random markers have important limitations: they are predominantly dominant (unable to distinguish heterozygotes), highly sensitive to reaction conditions, and frequently exhibit poor reproducibility across laboratories. More importantly, because they are typically linked to rather than embedded within the sex-determining region (SDR), recombination can decouple the marker from the sex-determining locus and reduce diagnostic reliability.

To improve reproducibility of random bands, researchers frequently cloned and sequenced these distinct fragments to convert them into longer, locus-specific Sequence

Characterized Amplified Region (SCAR) markers, as successfully deployed for early sex screening in *Phoenix dactylifera* L. [41] and *Hippophae rhamnoides* L. [42]. Co-dominant and highly polymorphic Simple Sequence Repeat (SSR) markers were widely utilized to map and track sex-linked alleles. Although SCAR and SSR markers improved assay specificity compared with random amplification methods, many remained correlative rather than mechanistic. Their diagnostic performance therefore depended on linkage distance, recombination rate, and the genetic background in which they were validated. Historical linked markers were thus important for early marker discovery, but they provided limited cross-population or cross-species transferability.

3.2. SDR-Linked SNPs and High-Throughput Genotyping

Next-Generation Sequencing shifted the paradigm from searching for random linked markers to directly identifying variants within or near the SDR. Techniques such as BSA-seq and Genome-Wide Association Studies (GWAS) are now standard methods for rapidly pinpointing sex-linked genomic intervals. For instance, by leveraging population-scale resequencing, researchers successfully delineated a precise 21-kb male-specific region (MSR) on the sex chromosome of *S. oleracea* [43]. Similarly, WGS of sex-segregating hybrid populations in *Actinidia arguta* (Siebold & Zucc.) Planch. ex Miq. enabled the rapid development of highly accurate sex-specific markers; however, their poor transferability to *Actinidia chinensis* Planch. highlights the highly lineage-specific nature of homomorphic SDR boundaries [44]. This sequencing-based localization differs from earlier workflows such as the RAPD-to-SCAR development of the papaya sex determination marker in *C. papaya*, which relied on linked fragments rather than direct resolution of SDR architecture [35].

Once sex-linked intervals are identified, diagnostic variants can be converted into scalable assays such as sequence-tagged site (STS), KASP, or SNP genotyping platforms. In *C. sativa*, large-scale InDel profiling has helped resolve population structure and identify stable sex-linked diagnostic assays within differentiated chromosomal regions [45]. In *A. officinalis*, early sex-linked primer sets have been upgraded into male-specific STS markers for rapid seedling screening [46,47]. Similarly, in *C. sativus*, BSA-seq mapping of the gynoecious trait enabled the development of KASP markers for sex-expression-related breeding traits [48]. These examples show that high-throughput genotyping platforms can improve scalability and standardization, but their reliability still depends on SDR stability and validation across genetic backgrounds.

3.3. Structural, PAV- and CNV-Based Markers

Long-read sequencing and pangenomic approaches have expanded sex-marker development beyond SNPs and small InDels to structural variants (SVs), presence/absence variations (PAVs), and copy number variations (CNVs). These variants are particularly relevant for sex identification because they can mark large Y- or W-linked regions, define sex-specific insertions or deletions, or quantify dosage differences associated with sex determination. Compared with linked SNPs, SV-, PAV-, and CNV-based markers can provide more direct access to structural or dosage-based diagnostic targets, especially in species with complex SDRs, dynamic sex chromosome turnover, or heteromorphic sex chromosomes.

Within the specific context of sex determination, recent long-read assemblies have revealed crucial gene rearrangements within the massive Y chromosome of *S. latifolia* [17]. In *A. palmeri*, a 1.3 Mb insertion constitutes the male-specific region [27]. In *Juglandaceae* species, a 4.5 Mb inversion dictates the protandrous or protogynous flowering type [49]. For large PAVs or inversions, marker development can focus on presence/absence assays or breakpoint-spanning junction PCR. In contrast, dosage-dependent architectures require

quantitative approaches such as qPCR, ddPCR, or read-depth-based CNV analysis. ddPCR can be particularly useful for dosage-sensitive targets because it enables absolute quantification without standard curves and can resolve subtle copy number differences [50]. Thus, SV-, PAV-, and CNV-based markers extend sex identification from sequence-level polymorphisms to structural and quantitative diagnostic targets.

3.4. Diagnostic Reliability of Linked, Functional and Structural Markers

The evolution of sex-marker technologies highlights a central principle: diagnostic reliability generally increases as markers move closer to the causal sex-determining target. In marker-assisted selection more broadly, marker performance depends strongly on the linkage relationship between the marker and the causal locus [51]. Historically, many sex-linked markers were correlative and lacked direct functional connection to sex-determining genes, making them vulnerable to recombination and population-specific haplotype structure [52].

Linked markers, including RAPD, AFLP, SSRs, SCARs, and some NGS-derived InDels, are statistically associated with sex but may be physically distant from the causal SDR. Their reliability can therefore decline when recombination, SDR turnover, or population divergence breaks the association between marker and trait. This problem is illustrated by recent sex-linked markers developed in *Hippophae tibetana* Schltdl. and *Hippophae salicifolia* D.Don, which showed high accuracy within target species but limited transferability across the genus [53,54]. A similar pattern was observed in *A. arguta*, where markers with high within-species accuracy were not transferable to *A. chinensis* [44].

Functional markers provide a higher level of diagnostic confidence when causal genes or functional variants are known. For example, markers targeting *FrBy* and *SyGI* in *Actinidia* can achieve high diagnostic accuracy in validated genetic backgrounds because they assay genes directly involved in sex determination [19,55]. Structural markers can be similarly stable when sex is governed by causal PAVs, inversions, Y/W-specific sequences, or dosage variation, because they track the underlying structural event rather than a distant linked polymorphism. Nevertheless, even functional or structural markers should not be assumed to be universally transferable. Cross-population and cross-lineage validation remain essential, especially in taxa with dynamic SDR turnover, high heterozygosity, polyploidy, or environmentally labile sex expression.

3.5. Marker Development in Complex Genomes

Despite advances in sequencing and marker design, reliable sex-marker development remains challenging in polyploid, highly heterozygous, and non-model species. Polyploidy complicates allele dosage estimation and genotypic classification, especially when sex expression depends on allele ratios or chromosome dosage rather than binary locus presence. In such cases, qualitative presence/absence assays may be insufficient, and dosage-aware genotyping becomes necessary [56]. High heterozygosity can also obscure sex-linked haplotypes and reduce mapping resolution, underscoring the need for haplotype-aware analysis in complex genomes [57].

Non-model dioecious species present additional challenges because they often lack high-quality reference genomes, well-characterized SDRs, or broad validation panels [58]. Moreover, lineage-specific gene duplications and paralogue interference can complicate primer design and lead to ambiguous amplification, particularly when sex-determining candidates arise from duplicated regulatory modules [59]. These constraints explain why markers developed from one reference genotype or mapping population may fail in broader germplasm. Reliable marker development in complex genomes therefore requires ploidy-aware genotyping, haplotype resolution, careful primer validation, and testing

across diverse genetic backgrounds. These challenges also motivate future efforts toward haplotype-resolved and graph-based pangenomes [60]. At commercial scale, these limitations can translate into diagnostic failure through three main routes. First, linkage-based assays may fail when recombination, SDR turnover, or population-specific haplotypes decouple the marker from the causal sex-determining locus. Second, genome complexity, including polyploidy, homoeologous or paralogous sequence interference, allele-dosage ambiguity, and possible gene silencing, can reduce assay specificity or produce ambiguous amplification patterns. Third, DNA markers identify inherited genetic sex, but they do not necessarily predict realized sex expression when sex is modulated by developmental stage, hormonal regulation, environmental stress, or sex reversal. Therefore, reliable application requires not only marker validation across diverse germplasm and production environments, but also a clear distinction between genetic sex and realized developmental, physiological, or phenotypic sex. This distinction provides the rationale for a dual-layer strategy in which molecular assays provide confirmatory evidence for genetic sex, whereas non-invasive phenotyping captures downstream physiological or phenotypic states.

4. Non-Invasive Phenotyping for Realized Sex Identification

Building on the diagnostic limitations discussed above, non-invasive phenotyping provides a complementary layer for assessing realized physiological or phenotypic sex expression. Unlike DNA markers, which infer inherited genetic sex, optical and sensor-based approaches detect downstream biochemical, structural, or physiological traits that may reflect sex-related developmental states. Therefore, these tools should not be viewed as replacements for validated molecular markers, but as prescreening or physiological-state assessment methods within a dual-layer identification strategy.

4.1. Genetic Sex and Realized Sex Expression

To understand the role of phenotyping in sex identification, plant sex should be conceptualized as a hierarchical process rather than a single static trait. Genetic sex is determined by chromosomes, SDRs, causal genes, structural variants, or dosage variation, whereas realized sex expression emerges through downstream developmental, physiological, and phenotypic layers. Developmental sex reflects gene expression and hormonal regulation during organogenesis. Downstream, physiological sex is represented by metabolic and biochemical states, and phenotypic sex corresponds to visible sexual expression [30]. This layered view, summarized in Figure 1B, explains why DNA markers and spectral phenotyping provide complementary rather than interchangeable information.

As discussed above, the limitation most relevant to non-invasive phenotyping is that DNA-based assays primarily detect genetic sex, whereas production outcomes often depend on realized sex expression under specific developmental or environmental conditions. As highlighted by phenomics frameworks, plants exhibit strong morphological and physiological plasticity, and capturing the realized phenome under fluctuating environments remains a challenge that genomics alone cannot fully resolve [61]. When sex expression is influenced by developmental stage, hormonal regulation, or environmental stress, genetic sex may diverge from realized physiological or phenotypic sex, a process broadly associated with environmental sex reversal [62]. Such divergence can affect functional sex ratios under changing environments [63]. For example, natural feminization in the industrial dioecious crop *Herpetospermum pedunculosum* (Ser.) C.B.Clarke has been reported to constrain large-scale cultivation despite the availability of genomic markers [64]. In this context, spectral phenotyping may complement DNA markers by capturing metabolic or physiological states associated with realized sex expression.

4.2. Spectral Sensing of Sex-Related Physiological Signatures

Unlike genomic profiling, spectral technologies probe the interaction between light and molecular bonds, providing a chemical fingerprint of the downstream metabolic phenotype.

Raman spectroscopy (RS) specifically detects the vibrational energy levels of molecular bonds (e.g., C=C, C-H). Because plant sex differentiation is often accompanied by the differential accumulation of steroids, phenolics, or specific proteins, these metabolites generate distinct Raman peaks. Recently, Holman et al. [65] successfully used a handheld Raman spectrometer to identify the sex of mature *A. palmeri* in the field. Their analysis revealed that spectral differences between male and female leaves primarily stemmed from fluctuations in nitrates, lipids, carotenoids, and terpenes.

To address the low sensitivity of conventional RS, researchers have introduced nanotechnology. Surface-Enhanced Raman Scattering (SERS) platforms excel at the ultrasensitive, non-destructive detection of trace effector molecules and secondary metabolites that dictate real-time physiological states [66]. Applying this nanotech-amplification to sex identification, Jiang et al. [55] successfully distinguished the sex of *Actinidia* spp. without nucleic acid amplification, reducing the detection time to mere minutes. SERS can improve the signal-to-noise ratio, enabling the detection of trace biochemicals in complex field environments.

Operating on a different physical principle, reflectance rather than scattering, hyperspectral imaging (HSI) has emerged as a powerful tool in high-throughput phenotyping (HTP), allowing researchers to capture both spatial and rich spectral information simultaneously across complex plant structures [67]. The rapid adoption of such aerial and ground-based HTP platforms equipped with multi-sensor arrays has reduced the phenotyping bottleneck in modern crop breeding [68]. Matros et al. [69] demonstrated that utilizing a high-resolution broadband spectroradiometer on *C. sativa* not only accurately distinguished dioecious individuals but also simultaneously monitored the nutritional status of the population, paving the way for high-throughput, contactless field assessments. Other non-invasive phenotyping platforms further illustrate the expanding technical basis for physiological-state monitoring. For example, handheld fluorescence imaging has been used to monitor photosynthetic activity and quantify anthocyanin accumulation non-destructively [70], while automated multi-sensor systems integrating RGB, depth, spectral fluorescence, and thermal sensing have enabled continuous monitoring of growth and physiological traits in plant propagation systems [71]. Although these approaches have not yet been widely applied to sex identification, they demonstrate how optical and multi-sensor phenotyping can capture pigment-related, photosynthetic, structural, and stress-associated traits that may contribute to sex-related physiological signatures. However, the deployment of HTP spectral tools generates high-dimensional sensor data; extracting meaningful physiological and sex-linked patterns from these datasets requires appropriate machine-learning or multivariate analytical approaches [72].

4.3. Machine-Learning-Assisted Spectral Classification

Given the high-dimensional and non-linear nature of spectral data, machine learning (ML) algorithms are increasingly used to decode complex spectral signatures in plant phenotyping. In sex identification, ML can support not only binary male–female classification but also the detection of intermediate or unstable sexual states. In *C. sativa*, Raman spectra analyzed with Support Vector Machines (SVMs) and Partial Least Squares Discriminant Analysis (PLS-DA) captured differences in carotenoid-associated bands at 1155 and 1525 cm^{-1}, enabling accurate seedling-stage sex classification in the tested material [16]. Similarly, Raman-based models have been used to distinguish unisexual plants

from hermaphrodites, a task that is difficult to resolve using DNA markers alone but relevant for removing unstable individuals in breeding populations [73].

However, high classification accuracy in small datasets should be interpreted cautiously. Spectral datasets often contain many more variables than biological samples, making models vulnerable to overfitting or artifact memorization. In such cases, algorithms may learn variation associated with leaf age, greenhouse background, water status, or nutrient conditions rather than genuine sex-linked biochemical features [74]. Therefore, spectral sex-identification models should be validated using independent populations, developmental stages, environments, and instruments. Model interpretability is equally important: feature attribution, saliency mapping, or wavelength-importance analysis can help determine whether the discriminative spectral regions correspond to plausible metabolites or chemical bonds associated with sex differentiation. Thus, ML should be viewed not only as a classifier, but also as a tool for linking spectral signals to biologically interpretable sex-related metabolic signatures.

4.4. Spectral Phenotyping as a Triage Strategy

Although these proof-of-concept studies demonstrate promising potential, spectral sex identification should currently not be viewed as a definitive replacement for molecular markers. Its main limitations include environmental background noise, instrument-dependent variation, limited field validation, and the sensitivity of portable devices. Therefore, its most realistic near-term role is high-throughput prescreening rather than stand-alone diagnosis, especially in large breeding populations where exhaustive molecular genotyping may be costly or time-consuming. This role is consistent with field-based high-throughput phenotyping frameworks designed to alleviate phenotyping bottlenecks in breeding programs [75].

Spectral phenotyping can therefore serve as a triage layer: models may prioritize high-confidence and ambiguous samples, while mechanism-appropriate molecular markers provide confirmatory genotyping. Related phenomic-selection studies suggest that non-destructive spectral data can complement genomic information in breeding workflows, although predictive performance remains context-dependent and requires validation across populations and environments [76,77]. For sex identification, this means that spectral models should be used to reduce and prioritize genotyping demand, not to replace molecular confirmation. Rigorous validation and careful control of biological and computational confounders are therefore essential (see Box 1).

Box 1. Key cautions for spectral and machine-learning-based sex identification.

1. Limited direct evidence
Direct applications remain restricted to a small number of species.
2. Small-sample overfitting
High-dimensional spectral data combined with small datasets may inflate classification accuracy.
3. Confounding factors
Leaf age, water status, nutrient level, stress, genotype and environment may dominate spectral variation.
4. Ground-truth uncertainty
Seedling sex labels require later flowering records or molecular confirmation.
5. Poor transferability
Models trained on one cultivar, environment, year, or instrument may not generalize.
6. Need for biochemical interpretation
Important wavelengths should be linked to metabolites or physiological processes rather than treated as uninterpreted model features.
Until these issues are addressed, spectral models should be treated as prescreening tools rather than stand-alone diagnostic assays.

5. A Mechanism-Guided Decision Framework

Building on the mechanism–target relationships summarized in Figure 1 and the diagnostic technologies reviewed in Sections 3 and 4, we propose a mechanism-guided decision framework for selecting early sex-identification strategies (Figure 3).

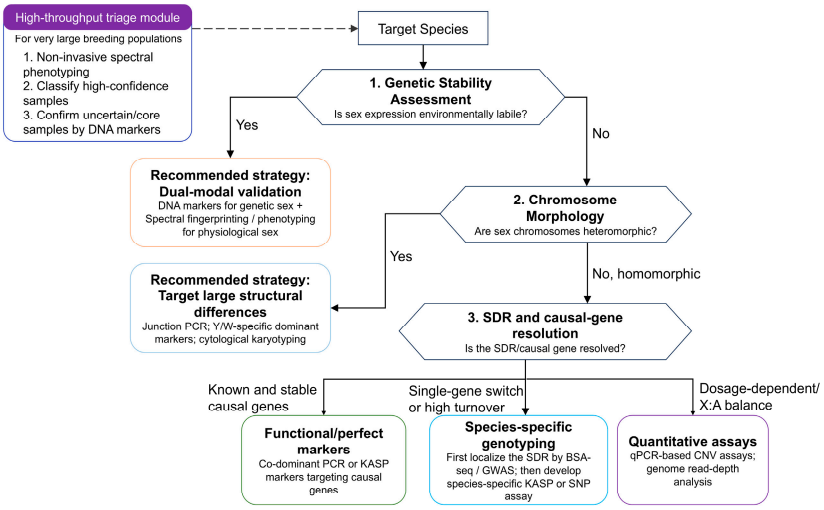

Figure 3. A mechanism-guided decision tree for selecting early sex-identification technologies in dioecious plants. The decision tree links sex-determining mechanisms with appropriate diagnostic strategies, including dual-modal validation, structural assays, functional markers, species-specific genotyping, and quantitative dosage assays. For large breeding populations, spectral phenotyping and machine learning can be used as a preliminary triage step before mechanism-appropriate molecular confirmation. Box colors are used to distinguish different recommendation modules or diagnostic strategies and do not indicate quantitative ranking; solid arrows indicate the main decision pathways, whereas the dashed arrow indicates an optional high-throughput triage step for very large breeding populations.

5.1. Matching Mechanisms, Targets and Technologies

In this framework, the sex determination mechanism defines the diagnostic target, which in turn guides the choice of detection technology. Available sex-identification tools differ substantially in reliability, scalability, cost, invasiveness, and validation requirements (Table 2). When these comparative metrics are considered together with the decision tree (Figure 3), technology selection can follow the underlying genetic architecture. Heteromorphic systems or large PAV-driven systems are best addressed by assays targeting Y/W-specific regions, structural breakpoints, or sex-linked insertions; stable homomorphic SDRs with known causal genes can be targeted using functional markers such as gene-specific PCR or KASP; and dynamic or dosage-dependent systems often require species-specific SDR mapping and quantitative assays such as qPCR, ddPCR, or read-depth-based CNV analysis. Representative studies illustrate how the technologies summarized in Table 2 can be matched to different diagnostic targets. For example, sex-specific regions, linked fragments, or large PAVs can support presence/absence PCR, SCAR markers, or structure-targeted PCR assays, as shown by marker development in *P. dactylifera*, *H. rhamnoides*, *A. palmeri*, and *C. sativa* [27,41,42,45]. Stable SDR-linked variants or candidate genes can be converted into gene-specific PCR, KASP, or SNP-based assays in systems such as *Actinidia*, *S. oleracea*, and *C. sativus* [19,20,43,44,48]. Dosage-dependent systems such as *H. lupulus* require quantitative approaches, including qPCR, ddPCR, or read-depth-based CNV analysis [28,29,51]. Non-invasive approaches, including Raman spectroscopy, SERS, and hyperspectral imaging, have been explored in *A. palmeri*, *Actinidia* spp., and *C. sativa* as prescreening tools for sex-related physiological signatures [16,55,65,69,73].

Table 2. Comparative overview of major technologies for early sex identification in dioecious plants.

Technology	Detection Layer	Reliability	Throughput	Per-Sample Cost	Invasiveness	Best Use Case	Main Limitation
Morphology	Vegetative/Floral phenotype	Low for vegetative traits/high at flowering	High	Very low	Non-invasive	Flowering-stage confirmation	Late detection; highly environment-sensitive
RAPD/AFLP/ISSR	Random DNA fragments	Low to moderate	Low	Low	Invasive	Preliminary marker discovery	Poor reproducibility; high recombination risk
SCAR/SSR	Linked or locus-specific markers	Moderate when validated	Moderate	Low	Invasive	Seedling screening in validated species	Limited cross-species transferability
KASP/SNP arrays	SDR-linked variants	High when SDR-linked and validated	Very high	Low–moderate	Invasive	Breeding-scale automated genotyping	Requires prior SDR mapping and validation
Junction PCR for SV/PAV	Structural variants/Insertions	High when breakpoint-specific	Moderate–high	Low	Invasive	PAV- or inversion-based sex diagnosis	Requires high-quality assemblies or known breakpoints
qPCR/ddPCR/read-depth CNV analysis	Copy number/dosage variation	High when dosage threshold is validated	Moderate	Low–moderate	Invasive	Dosage-dependent sex systems	Requires quantitative calibration and appropriate controls
Raman/SERS + ML	Metabolic fingerprint	Emerging/Context-dependent	High	Low after platform setup	Non-invasive	Rapid prescreening/physiological-state validation	Model overfitting; limited cross-environment validation
Hyperspectral imaging + ML	Canopy/leaf optical traits	Emerging/Context-dependent	Very high	Low after platform setup	Non-invasive	Large-scale field prescreening/canopy-level monitoring	Strong environmental and background confounding

Note: Reliability, throughput, cost, and invasiveness are relative assessments based on typical use cases rather than universal performance values. For DNA-based methods, reliability depends on linkage distance, causal relevance, breakpoint specificity, dosage calibration, and validation across genetic backgrounds. Spectral and machine-learning-based methods are classified as context-dependent because their direct application to plant sex identification remains limited and requires broader validation across genotypes, developmental stages, environments, and instruments.

5.2. Spectral Prescreening and Molecular Confirmation

While molecular markers provide the most reliable basis for confirmatory genotyping, applying DNA extraction and KASP genotyping to thousands of seedlings can remain costly and time-consuming. For large breeding populations, spectral phenotyping may therefore serve as a preliminary triage layer before molecular confirmation (Figure 3, triage module). Non-invasive optical tools, such as Raman spectroscopy or hyperspectral sensing, combined with machine-learning models, can be used to prioritize high-confidence, uncertain, or physiologically stressed individuals for subsequent analysis.

In this two-step workflow, spectral models are used for rapid prescreening, whereas DNA-based assays provide mechanism-appropriate confirmation. Individuals in the uncertainty zone, or those of particular breeding value, can be genotyped using KASP, junction PCR, qPCR, ddPCR, or read-depth-based CNV analysis, depending on the relevant diagnostic target. This strategy does not treat spectral phenotyping as a substitute for molecular markers, but as a front-end filtering tool that can reduce unnecessary genotyping and improve sampling efficiency. Its practical value depends on validation across genotypes, developmental stages, environments, and instruments.

5.3. Context-Specific Deployment of Sex-Identification Strategies

The application of a mechanism-guided framework depends on the operational objective, population size, and acceptable level of diagnostic uncertainty. In large breeding populations, where thousands of seedlings may need to be screened before flowering, throughput and cost efficiency are often the primary considerations. In this context, spectral prescreening can be used to prioritize individuals for subsequent molecular confirmation, especially when the goal is to reduce unnecessary genotyping while retaining a reliable confirmatory layer.

In germplasm collections and parental-line selection, by contrast, diagnostic accuracy and long-term genetic traceability are more important than screening speed. Here, causal or tightly linked molecular markers, such as functional SDR markers, KASP assays, or structural markers, should be prioritized and validated across the relevant genetic backgrounds. For field populations exposed to environmental stresses, genetic sex information alone may be insufficient when sex expression is developmentally or environmentally labile. In such cases, combining molecular markers with physiological or spectral phenotyping can help distinguish underlying genetic sex from realized sex expression.

Overall, the decision framework should not be interpreted as a fixed pipeline, but as a context-dependent strategy for matching diagnostic tools to biological mechanisms and practical objectives. By aligning sex-determining architecture with appropriate genetic, structural, dosage-based, or phenotyping tools, researchers and breeders can develop more strong and species-specific screening pipelines for dioecious plant breeding, cultivation, and germplasm management.

6. Challenges and Future Perspectives

Despite recent progress in sex-identification technologies, translating laboratory proof-of-concept assays into robust field applications still requires advances in marker validation, mechanistic confirmation, and scalable deployment. Future research should address the reliability of molecular markers, the functional validation of sex-determining candidates, the resolution of complex SDRs, the robustness of spectral models, and the integration of genetic, physiological, and environmental information.

6.1. Cross-Population Marker Validation

A major challenge in molecular diagnostics is ascertainment bias, which occurs when markers are developed from a narrow reference panel or specific genetic background and then perform poorly in broader germplasm [78]. This issue is particularly important for sex identification because many markers are discovered in bi-parental populations, limited geographical cohorts, or species with rapid SDR turnover and high heterozygosity. Future studies should therefore prioritize cross-population and cross-lineage validation of newly developed markers across diverse accessions and related taxa [79]. For practical sex identification, marker performance should be evaluated in the target breeding or conservation context rather than assumed from the discovery population alone.

6.2. Functional Validation of Sex Determination Candidates

High-throughput mapping approaches, such as GWAS and BSA-seq, can localize SDRs, but correlation does not establish causation. Future studies should move from locus association toward functional validation of candidate sex-determining genes. Reverse-genetic tools, including CRISPR/Cas9-mediated mutagenesis and virus-induced gene silencing, provide useful approaches for testing candidate functions, as illustrated by the development of VIGS protocols in *S. latifolia* and the broader application of genome editing in reproductive traits [80,81]. Functional perturbation of male-promoting or female-suppressing candidates can determine whether a sex-linked locus is causal rather than merely associated with sex expression. For sex identification, such validation helps ensure that diagnostic assays target true functional determinants rather than transiently linked regions.

6.3. Haplotype-Resolved Pangenomes for Complex SDRs

The repetitive and structurally complex nature of sex chromosomes means that a single linear reference genome is often insufficient. A single reference individual cannot capture the full spectrum of PAVs, CNVs, and haplotype-specific sequences within a species [82]. This limitation is particularly relevant in dioecious species, where Y- or W-specific regions may be collapsed, misassembled, or omitted from reference assemblies because of limited standards for representing sex chromosomes [83]. In lineages with frequent sex chromosome turnover, such as *Salix*, single-reference assemblies may also miss lineage-specific SDR rearrangements and newly evolved sex-linked structural variants [26]. Therefore, haplotype-resolved and graph-based pangenomes provide an important future direction for sex-determination genetics [84]. For sex identification, these resources will be valuable for discovering SVs, PAVs, CNVs, and haplotype-specific markers that are absent from single-reference genomes.

6.4. Multi-Environment Validation of Spectral Models

While Raman and hyperspectral sensing provide promising routes for non-invasive sex screening, their practical use depends on model robustness across environments. A spectral model trained under controlled greenhouse conditions may fail in the field because light intensity, leaf age, water status, nutrient availability, and stress history can all influence optical or metabolic signals. Future studies should therefore adopt multi-environment validation frameworks, using independent datasets collected across years, genotypes, developmental stages, instruments, and cultivation conditions. The goal is to distinguish genuine sex-linked metabolic fingerprints from environmental noise, rather than simply detecting drought, heat, or nutrient stress. Advances in artificial intelligence may help model complex genotype–environment–phenotype relationships [85], but predictive performance should always be evaluated with independent validation sets rather than internal cross-validation alone. Spectral signatures will be more credible when they consistently

predict physiological sex across diverse field conditions and are supported by interpretable biochemical features.

6.5. Multimodal Decision Support for Sex Identification

Future sex-identification pipelines will likely require the integration of multiple information layers rather than reliance on a single diagnostic method. DNA markers provide information on genetic sex, spectral phenotyping captures physiological or metabolic states, and environmental metadata help interpret when sex expression may become labile. Multi-omics integration has already become an increasingly useful strategy for dissecting complex plant traits [86], and similar logic can be applied to dioecious plant breeding. For sex identification, multimodal decision-support systems could combine SDR markers, spectral fingerprints, and environmental variables to estimate both baseline genetic sex and the risk of environmentally induced sex reversal [86,87]. Such systems should be developed as practical, species-specific tools for breeding and germplasm management rather than as generic AI platforms. By linking genomic precision with physiological and environmental context, multimodal integration can improve the robustness of early sex screening in environmentally sensitive dioecious crops.

7. Conclusions

Early sex identification remains a major prerequisite for efficient breeding and cultivation of dioecious plants. This review highlights that the choice of diagnostic technology should be guided by the underlying sex-determining mechanism rather than by methodological convenience alone. Heteromorphic sex chromosomes, homomorphic SDRs, functional sex-determining genes, sex chromosome turnover, structural variants, dosage-dependent systems, and environmentally labile sex expression each generate distinct diagnostic targets and therefore require different marker-development strategies.

Molecular markers remain the most reliable tools for confirmatory sex identification, especially when they target causal genes, SDR-linked variants, structural breakpoints, or dosage differences. However, DNA-based assays primarily detect genetic sex and may not fully capture developmentally or environmentally modulated physiological sex. Emerging spectral and machine-learning approaches provide a promising but still exploratory route for non-invasive prescreening, particularly in large breeding populations or stress-prone field environments.

By combining genomic information with phenotypic adaptability, a mechanism-guided framework can support more reliable, scalable, and species-specific sex-identification pipelines. Future progress will depend on cross-population marker validation, functional confirmation of causal genes, haplotype-resolved SDR assemblies, multi-environment validation of spectral models, and the development of multimodal decision-support systems that can improve propagation efficiency, planting-material quality, and sustainable cultivation of dioecious horticultural plants.

Author Contributions: Conceptualization, J.Z.; methodology, J.Z.; software, H.W.; validation, H.W.; formal analysis, J.Z.; investigation, K.S.; resources, H.S.; data curation, Y.Q.; writing—original draft preparation, K.S.; writing—review and editing, H.S.; visualization, A.L.; supervision, J.L. and A.L.; project administration, J.L.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Nonprofit Research Institution of the Chinese Academy of Forestry, grant number CAFYBB2023MA003.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This work is dedicated to the memory of our esteemed and beloved supervisor, Zhenyuan Sun. He was an exceptional and rigorous plant physiologist whose guidance and spirit deeply inspired this research. He is deeply missed. Some icons in Figures 1 and 2 were adapted from resources created by Flaticon contributors and used in accordance with the Flaticon free license.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. Author Ke Shi was employed by the company Beijing Jindu Landscape and Afforesting Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFLP	Amplified fragment length polymorphism
AI	Artificial intelligence
BSA	Bulked segregant analysis
CNV	Copy number variation
ddPCR	Droplet digital PCR
G × E	Genotype-by-environment interaction
GWAS	Genome-wide association study
HSI	Hyperspectral imaging
HTP	High-throughput phenotyping
InDel	Insertion–deletion
ISSR	Inter-simple sequence repeat
KASP	Kompetitive allele-specific PCR
ML	Machine learning
MSR	Male-specific region
NGS	Next-generation sequencing
PAV	Presence/absence variation
PCR	Polymerase chain reaction
PLS-DA	Partial least squares discriminant analysis
qPCR	Quantitative PCR
RAPD	Random amplified polymorphic DNA
RGB	Red–green–blue
RS	Raman spectroscopy
SCAR	Sequence-characterized amplified region
SDR	Sex-determining region
SERS	Surface-enhanced Raman scattering
siRNA	Small interfering RNA
SNP	Single-nucleotide polymorphism
SSR	Simple sequence repeat
STS	Sequence-tagged site
SV	Structural variant
SVM	Support vector machine
VIGS	Virus-induced gene silencing
WGS	Whole-genome sequencing
X:A	X-chromosome-to-autosome ratio

References

1. Charlesworth, D.; Harkess, A. Why Should We Study Plant Sex Chromosomes? *Plant Cell* **2024**, *36*, 1242–1256. [CrossRef] [PubMed]
2. Charlesworth, D. Plant Sex Chromosomes. *Annu. Rev. Plant Biol.* **2016**, *67*, 397–420. [CrossRef] [PubMed]

3. Charlesworth, D. Plant Sex Chromosome Evolution. *J. Exp. Bot.* **2013**, *64*, 405–420. [CrossRef] [PubMed]
4. Renner, S.S.; Müller, N.A. Sex Determination and Sex Chromosome Evolution in Land Plants. *Phil. Trans. R. Soc. B* **2022**, *377*, 20210210. [CrossRef] [PubMed]
5. Razumova, O.V.; Alexandrov, O.S.; Bone, K.D.; Karlov, G.I.; Divashuk, M.G. Sex Chromosomes and Sex Determination in Dioecious Agricultural Plants. *Agronomy* **2023**, *13*, 540. [CrossRef]
6. Kazama, Y.; Kobayashi, T.; Filatov, D.A. Evolution of Sex-Determination in Dioecious Plants: From Active Y to X/A Balance? *BioEssays* **2023**, *45*, 2300111. [CrossRef] [PubMed]
7. Henry, I.M.; Akagi, T.; Tao, R.; Comai, L. One Hundred Ways to Invent the Sexes: Theoretical and Observed Paths to Dioecy in Plants. *Annu. Rev. Plant Biol.* **2018**, *69*, 553–575. [CrossRef] [PubMed]
8. Han, L.; Li, M.; Li, C.; Zhao, B.; Wang, Z.; Liu, Y.; Liu, Z.; Huang, Y.; Liu, L.; Geng, H.; et al. ARF3-Mediated Auxin Signaling Is Essential for Sex Determination in Cucumber. *Science* **2025**, *391*, 59–63. [CrossRef] [PubMed]
9. He, L.; Jia, K.-H.; Zhang, R.-G.; Wang, Y.; Shi, T.-L.; Li, Z.-C.; Zeng, S.-W.; Cai, X.-J.; Wagner, N.D.; Hörandl, E.; et al. Chromosome-Scale Assembly of the Genome of *Salix dunnii* Reveals a Male-Heterogametic Sex Determination System on Chromosome 7. *Mol. Ecol. Resour.* **2021**, *21*, 1966–1982. [CrossRef] [PubMed]
10. Hyden, B.; Carper, D.L.; Abraham, P.E.; Yuan, G.; Yao, T.; Baumgart, L.; Zhang, Y.; Chen, C.; O'Malley, R.; Chen, J.-G.; et al. Functional Analysis of Salix Purpurea Genes Support Roles for ARR17 and GATA15 as Master Regulators of Sex Determination. *Plant Direct* **2023**, *7*, e546. [CrossRef] [PubMed]
11. Li, C.; Dong, S.; Liu, X.; Guan, J.; Beckles, D.M.; Gu, X.; Sun, J.; Miao, H.; Zhang, S. Co-Domestication of Cold Tolerance and Female Flower Is Determined by *CsEIN2* in Cucumber. *Plant Biotechnol. J.* **2025**, *23*, 4412–4427. [CrossRef] [PubMed]
12. You, C.; Yang, H.; Zhao, Y.; Wang, X.; Wei, S.; Chen, N.; Zhang, Y.; Liu, L.; Qian, W.; Li, S.; et al. Spatiotemporal Transcriptomic Atlas Reveals the Regulatory Mechanisms Underlying Early Inflorescence Development and Sex Differentiation in *Spinach*. *Adv. Sci.* **2025**, *12*, e07818. [CrossRef] [PubMed]
13. Shi, J.; Toscani, M.; Dowling, C.A.; Schilling, S.; Melzer, R. Identification of Genes Associated with Sex Expression and Sex Determination in Hemp (*Cannabis sativa* L.). *J. Exp. Bot.* **2025**, *76*, 175–190. [CrossRef] [PubMed]
14. Prentout, D.; El Aoudati, S.; Mathis, F.; Marais, G.A.B.; Henri, H. Promising High Fidelity Genetic Markers for Sexing *Cannabis sativa* Seedlings. *G3 Genes Genomes Genet.* **2025**, *15*, jkaf077. [CrossRef] [PubMed]
15. Zhou, X.; Zhang, M.; Zhang, L.; Hu, J. GWAS-Assisted Genomic Selection for Achieving High-Precision Early Sex Prediction in *Populus deltoides*. *Plant Genome* **2026**, *19*, e70175. [CrossRef] [PubMed]
16. Higgins, S.; Jessup, R.; Kurouski, D. Raman Spectroscopy Enables Highly Accurate Differentiation between Young Male and Female Hemp Plants. *Planta* **2022**, *255*, 85. [CrossRef] [PubMed]
17. Moraga, C.; Branco, C.; Rougemont, Q.; Jedlička, P.; Mendoza-Galindo, E.; Veltsos, P.; Hanique, M.; Rodríguez de la Vega, R.C.; Tannier, E.; Liu, X.; et al. The *Silene latifolia* Genome and Its Giant Y Chromosome. *Science* **2025**, *387*, 630–636. [CrossRef] [PubMed]
18. Johns, J.; Moore, M.L.; Escalona, M.; Miller, C.; Chumchim, N.; Nguyen, O.; Marimuthu, M.P.A.; Fairbairn, C.W.; Beraut, E.; Seligmann, W.E.; et al. A Chromosome Level Genome Assembly of the Marine Flowering Plant, Torrey's Surfgrass (*Phyllospadix torreyi*) Reveals an Exceptionally Large Y-Chromosome. *J. Hered.* **2025**, *117*, 566–576. [CrossRef] [PubMed]
19. Akagi, T.; Pilkington, S.M.; Varkonyi-Gasic, E.; Henry, I.M.; Sugano, S.S.; Sonoda, M.; Firl, A.; McNeilage, M.A.; Douglas, M.J.; Wang, T.; et al. Two Y-Chromosome-Encoded Genes Determine Sex in Kiwifruit. *Nat. Plants* **2019**, *5*, 801–809. [CrossRef] [PubMed]
20. Akagi, T.; Henry, I.M.; Ohtani, H.; Morimoto, T.; Beppu, K.; Kataoka, I.; Tao, R. A Y-Encoded Suppressor of Feminization Arose via Lineage-Specific Duplication of a Cytokinin Response Regulator in Kiwifruit. *Plant Cell* **2018**, *30*, 780–795. [CrossRef] [PubMed]
21. Akagi, T.; Varkonyi-Gasic, E.; Shirasawa, K.; Catanach, A.; Henry, I.M.; Mertten, D.; Datson, P.; Masuda, K.; Fujita, N.; Kuwada, E.; et al. Recurrent Neo-Sex Chromosome Evolution in Kiwifruit. *Nat. Plants* **2023**, *9*, 393–402. [CrossRef] [PubMed]
22. Zhao, J.; Wang, D.; Chen, K.; Xie, J.; Li, Y.; Wang, Y.; Zhao, L.; Yang, Y.; Olson, M.S.; Müller, N.A.; et al. A New Two-Gene System of Sex Determination in a Salicaceae. *Curr. Biol.* **2026**, *36*, 492–505.e5. [CrossRef] [PubMed]
23. Akagi, T.; Henry, I.M.; Tao, R.; Comai, L. A Y-Chromosome-Encoded Small RNA Acts as a Sex Determinant in Persimmons. *Science* **2014**, *346*, 646–650. [CrossRef] [PubMed]
24. Akagi, T.; Henry, I.M.; Kawai, T.; Comai, L.; Tao, R. Epigenetic Regulation of the Sex Determination Gene *MeGI* in Polyploid Persimmon. *Plant Cell* **2016**, *28*, 2905–2915. [CrossRef] [PubMed]
25. Renner, S.S.; Müller, N.A. Plant Sex Chromosomes Defy Evolutionary Models of Expanding Recombination Suppression and Genetic Degeneration. *Nat. Plants* **2021**, *7*, 392–402. [CrossRef] [PubMed]
26. Hu, N.; Sanderson, B.J.; Guo, M.; Feng, G.; Gambhir, D.; Hale, H.; Wang, D.; Hyden, B.; Liu, J.; Smart, L.B.; et al. Evolution of a ZW Sex Chromosome System in Willows. *Nat. Commun.* **2023**, *14*, 7144. [CrossRef] [PubMed]
27. Raiyemo, D.A.; Montgomery, J.S.; Cutti, L.; Abdollahi, F.; Llaca, V.; Fengler, K.; Lopez, A.J.; Morran, S.; Saski, C.A.; Nelson, D.R.; et al. Chromosome-level Assemblies of *Amaranthus palmeri*, *Amaranthus retroflexus*, and *Amaranthus hybridus* Allow for Genomic Comparisons and Identification of a Sex-determining Region. *Plant J.* **2025**, *121*, e70027. [CrossRef] [PubMed]

28. Akagi, T.; Segawa, T.; Uchida, R.; Tanaka, H.; Shirasawa, K.; Yamagishi, N.; Yaegashi, H.; Natsume, S.; Takagi, H.; Abe, A.; et al. Evolution and Functioning of an X–A Balance Sex-Determining System in Hops. *Nat. Plants* **2025**, *11*, 1339–1352. [CrossRef] [PubMed]
29. Dowling, C.A.; Michael, T.P.; McCabe, P.F.; Schilling, S.; Melzer, R. FT-like Genes in Cannabis and Hops: Sex Specific Expression and Copy-Number Variation May Explain Flowering Time Variation. *BMC Genom.* **2025**, *26*, 930. [CrossRef] [PubMed]
30. Zhao, W.; Zhu, K.; Ma, J.; He, Z.; Nie, L. The Mechanistic Insights into Sex Determination of Melon: Integrating Environmental Factors, Hormones and Genetic Regulation. *Plant Sci.* **2026**, *364*, 112999. [CrossRef] [PubMed]
31. Monthony, A.S.; Roy, J.; de Ronne, M.; Carlson, O.; Murch, S.J.; Torkamaneh, D. Sex-Specific Ethylene Responses Drive Floral Sexual Plasticity in *Cannabis sativa*. *Plant J.* **2026**, *125*, e70721. [CrossRef] [PubMed]
32. Ávila-Hernández, J.G.; Cárdenas-Aquino, M.d.R.; Camas-Reyes, A.; Martínez-Antonio, A. Sex Determination in Papaya: Current Status and Perspectives. *Plant Sci.* **2023**, *335*, 111814. [CrossRef] [PubMed]
33. Hyden, B.; Zou, J.; Wilkerson, D.G.; Carlson, C.H.; Robles, A.R.; DiFazio, S.P.; Smart, L.B. Structural Variation of a Sex-linked Region Confers Monoecy and Implicates *GATA15* as a Master Regulator of Sex in *Salix purpurea*. *New Phytol.* **2023**, *238*, 2512–2523. [CrossRef] [PubMed]
34. Baek, Y.; Vergara, D. A Review of Sexual Strategies in *Cannabis sativa* L. under Genomic and Environmental Controls. *Agrosystems Geosci. Environ.* **2025**, *8*, e70050. [CrossRef]
35. Urasaki, N.; Tokumoto, M.; Tarora, K.; Ban, Y.; Kayano, T.; Tanaka, H.; Oku, H.; Chinen, I.; Terauchi, R. A Male and Hermaphrodite Specific RAPD Marker for Papaya (*Carica papaya* L.). *Theor. Appl. Genet.* **2002**, *104*, 281–285. [CrossRef] [PubMed]
36. Ajuru, M.G. Sex Determination of *Carica papaya* L. Varieties at Seedling-Juvenile Stage Using Morpoho-Anatomical Characters. *J. Biol. Nat.* **2015**, *4*, 132–140.
37. Rezaei, M.; Rohani, A.; Heidari, P.; Lawson, S. Using Soft Computing and Leaf Dimensions to Determine Sex in Immature *Pistacia vera* Genotypes. *Measurement* **2021**, *174*, 108988. [CrossRef]
38. Fu, L.; Wang, Q.; Zhang, M.; Zheng, Y.; Wu, M.; Lan, Z.; Pu, J.; Zhang, H.; Chen, F.; Su, W.; et al. Electrochemical Sex Determination of Dioecious Plants Using Polydopamine-Functionalized Graphene Sheets. *Front. Chem.* **2020**, *8*, 92. [CrossRef] [PubMed]
39. Prakash, S.; Van Staden, J. Sex Identification in *Encephalartos natalensis* (Dyer and Verdoorn) Using RAPD Markers. *Euphytica* **2006**, *152*, 197–200. [CrossRef]
40. Goyat, S.; Grewal, A.; Singh, D.; Katiyar, R.S.; Tewari, S.K.; Nainwal, R.C.; Bindu, K.H. Sex-Linked AFLP Marker Identification in Dioecious Betelvine (*Piper betle* L.). *J. Hortic. Sci. Biotechnol.* **2019**, *94*, 422–427. [CrossRef]
41. Al-Qurainy, F.; Al-Ameri, A.A.; Khan, S.; Nadeem, M.; Gaafar, A.-R.Z.; Tarroum, M. SCAR Marker for Gender Identification in Date Palm (*Phoenix dactylifera* L.) at the Seedling Stage. *Int. J. Genom.* **2018**, *2018*, 3035406. [CrossRef] [PubMed]
42. Das, K.; Ganie, S.H.; Mangla, Y.; Dar, T.-U.-H.; Chaudhary, M.; Thakur, R.K.; Tandon, R.; Raina, S.N.; Goel, S. ISSR Markers for Gender Identification and Genetic Diagnosis of *Hippophae rhamnoides* ssp. Turkestanica Growing at High Altitudes in Ladakh Region (Jammu and Kashmir). *Protoplasma* **2017**, *254*, 1063–1077. [CrossRef] [PubMed]
43. She, H.; Xu, Z.; Zhang, H.; Li, G.; Wu, J.; Wang, X.; Li, Y.; Liu, Z.; Qian, W. Identification of a Male-Specific Region (MSR) in *Spinacia oleracea*. *Hortic. Plant J.* **2021**, *7*, 341–346. [CrossRef]
44. Guo, D.; Wang, R.; Fang, J.; Zhong, Y.; Qi, X. Development of Sex-Linked Markers for Gender Identification of *Actinidia arguta*. *Sci. Rep.* **2023**, *13*, 12780. [CrossRef] [PubMed]
45. Pan, G.; Li, Z.; Huang, S.; Tao, J.; Shi, Y.; Chen, A.; Li, J.; Tang, H.; Chang, L.; Deng, Y.; et al. Genome-Wide Development of Insertion-Deletion (InDel) Markers for Cannabis and Its Uses in Genetic Structure Analysis of Chinese Germplasm and Sex-Linked Marker Identification. *BMC Genom.* **2021**, *22*, 595. [CrossRef] [PubMed]
46. Ahmad, N.; Tian, R.; Li, G.; Zhao, C.; Fan, S.; Sun, J.; Zhao, S.; Wang, X. Establishment of Male-Specific Sequence-Tagged Site Markers in *Asparagus officinalis*: An Efficient Tool for Sex Identification. *Plant Breed.* **2022**, *141*, 471–481. [CrossRef]
47. Nakayama, H.; Ito, T.; Hayashi, Y.; Sonoda, T.; Fukuda, T.; Ochiai, T.; Kameya, T.; Kanno, A. Development of Sex-Linked Primers in Garden Asparagus (*Asparagus officinalis* L.). *Breed. Sci.* **2006**, *56*, 327–330. [CrossRef]
48. Sharma, E.; Dhall, R.K.; Verma, N.; Manchanda, P.; Bhatia, D.; Kumari, P.; Rana, N. Development of KASP Marker Associated with Gynoecious Trait Using BSA-Seq in *Cucumis sativus* L. *Physiol. Mol. Biol. Plants* **2025**, *31*, 1947–1961. [CrossRef] [PubMed]
49. Groh, J.S.; Vik, D.C.; Davis, M.; Monroe, J.G.; Stevens, K.A.; Brown, P.J.; Langley, C.H.; Coop, G. Ancient Structural Variants Control Sex-Specific Flowering Time Morphs in Walnuts and Hickories. *Science* **2025**, *387*, eado5578. [CrossRef] [PubMed]
50. He, Q.; Shang, X.; Tian, R.; Zhu, X.; Guo, W. An Efficient and Accurate Droplet Digital PCR Method for Rapid Transgene Copy Number Detection and Homozygous Identification in Cotton (*Gossypium hirsutum*). *Ind. Crops Prod.* **2023**, *204*, 117284. [CrossRef]
51. Song, L.; Wang, R.; Yang, X.; Zhang, A.; Liu, D. Molecular Markers and Their Applications in Marker-Assisted Selection (MAS) in Bread Wheat (*Triticum aestivum* L.). *Agriculture* **2023**, *13*, 642. [CrossRef]
52. Heikrujam, M.; Sharma, K.; Prasad, M.; Agrawal, V. Review on Different Mechanisms of Sex Determination and Sex-Linked Molecular Markers in Dioecious Crops: A Current Update. *Euphytica* **2015**, *201*, 161–194. [CrossRef]

53. Zeng, Z.; Zhang, S.; Tan, X.; Tso, N.; Shang, Z.; Zhang, J.; Li, W.; Wang, J.; Zhang, W.; Qiong, L. Development and Application of Sex-Specific Indel Markers for *Hippophae salicifolia* Based on Third-Generation Sequencing. *BMC Plant Biol.* **2025**, *25*, 692. [CrossRef] [PubMed]
54. Zeng, Z.; Wang, R.; Wang, J.; Chen, Y.; Wang, Y.; Song, Z.; Zhang, W.; Qiong, L. Development and Validation of Sex-Linked Molecular Markers for Rapid and Accurate Identification of Male and Female *Hippophae tibetana* Plants. *Sci. Rep.* **2024**, *14*, 19243. [CrossRef] [PubMed]
55. Jiang, H.; Zhu, H.; Yu, T.; Song, W.; Zhou, B.; Qu, C.; Su, M.; Liu, Y.; Miao, M.; Liu, H. Non-Amplification on-Spot Identifying the Sex of Dioecious Kiwi Plants by a Portable Raman Device. *Talanta* **2023**, *258*, 124447. [CrossRef] [PubMed]
56. Batista, L.G.; Mello, V.H.; Souza, A.P.; Margarido, G.R.A. Genomic Prediction with Allele Dosage Information in Highly Polyploid Species. *Theor. Appl. Genet.* **2022**, *135*, 723–739. [CrossRef] [PubMed]
57. Yuan, Y.; Scheben, A.; Edwards, D.; Chan, T.-F. Toward Haplotype Studies in Polyploid Plants to Assist Breeding. *Mol. Plant* **2021**, *14*, 1969–1972. [CrossRef] [PubMed]
58. Ungaro, A.; Pech, N.; Martin, J.-F.; McCairns, R.J.S.; Mévy, J.-P.; Chappaz, R.; Gilles, A. Challenges and Advances for Transcriptome Assembly in Non-Model Species. *PLoS ONE* **2017**, *12*, e0185020. [CrossRef] [PubMed]
59. Kaltenegger, E.; Ober, D. Paralogue Interference Affects the Dynamics after Gene Duplication. *Trends Plant Sci.* **2015**, *20*, 814–821. [CrossRef] [PubMed]
60. Jonkheer, E.M.; de Ridder, D.; van der Lee, T.A.J.; de Haan, J.R.; Berke, L.; Smit, S. Exploring Intra- and Intergenomic Variation in Haplotype-Resolved Pangenomes. *Plant Biotechnol. J.* **2025**, *23*, 874–886. [CrossRef] [PubMed]
61. Tardieu, F.; Cabrera-Bosquet, L.; Pridmore, T.; Bennett, M. Plant Phenomics, From Sensors to Knowledge. *Curr. Biol.* **2017**, *27*, R770–R783. [CrossRef] [PubMed]
62. Stelkens, R.B.; Wedekind, C. Environmental Sex Reversal, Trojan Sex Genes, and Sex Ratio Adjustment: Conditions and Population Consequences. *Mol. Ecol.* **2010**, *19*, 627–646. [CrossRef] [PubMed]
63. Edmands, S. Sex Ratios in a Warming World: Thermal Effects on Sex-Biased Survival, Sex Determination, and Sex Reversal. *J. Hered.* **2021**, *112*, 155–164. [CrossRef] [PubMed]
64. Li, W.-L.; Lu, Y.-J.; Ren, D.-S.; Li, J.-H.; Ma, Y.-B.; Li, X.-Y.; Li, M.-J.; Li, C.-X.; Wang, Y.; Ma, X.-R. The Sex-Determining Region and Early-Stage Gender Identification in the Dioecious Industrial Plant *Herpetospermum pedunculosum*. *Ind. Crops Prod.* **2025**, *236*, 122038. [CrossRef]
65. Holman, A.P.; Goff, N.K.; Juárez, I.D.; Higgins, S.; Rodriguez, A.; Bagavathiannan, M.; Kurouski, D.; Subramanian, N. Elucidation of Sex from Mature Palmer Amaranth (*Amaranthus palmeri*) Leaves Using a Portable Raman Spectrometer. *RSC Adv.* **2024**, *14*, 1833–1837. [CrossRef] [PubMed]
66. Du, S.; Yu, C.; Tang, L.; Lu, L. Applications of SERS in the Detection of Stress-Related Substances. *Nanomaterials* **2018**, *8*, 757. [CrossRef] [PubMed]
67. Sarić, R.; Nguyen, V.D.; Burge, T.; Berkowitz, O.; Trtílek, M.; Whelan, J.; Lewsey, M.G.; Čustović, E. Applications of Hyperspectral Imaging in Plant Phenotyping. *Trends Plant Sci.* **2022**, *27*, 301–315. [CrossRef] [PubMed]
68. Gill, T.; Gill, S.K.; Saini, D.K.; Chopra, Y.; de Koff, J.P.; Sandhu, K.S. A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. *Phenomics* **2022**, *2*, 156–183. [CrossRef] [PubMed]
69. Matros, A.; Menz, P.; Gill, A.R.; Santoscoy, A.; Dawson, T.; Seiffert, U.; Burton, R.A. Non-Invasive Assessment of Cultivar and Sex of *Cannabis sativa* L. by Means of Hyperspectral Measurement. *Plant-Environ. Interact.* **2023**, *4*, 258–274. [CrossRef] [PubMed]
70. Zhang, R.; Koh, S.S.; Teo, M.J.T.; Bi, R.; Zhang, S.; Dev, K.; Urano, D.; Dinish, U.S.; Olivo, M. Handheld Multifunctional Fluorescence Imager for Non-Invasive Plant Phenotyping. *Front. Plant Sci.* **2022**, *13*, 822634. [CrossRef] [PubMed]
71. Perdomo, S.A.; De la Paz, E.; Del Caño, R.; Seker, S.; Saha, T.; Wang, J.; Jaramillo-Botero, A. Non-Invasive in-Vivo Glucose-Based Stress Monitoring in Plants. *Biosens. Bioelectron.* **2023**, *231*, 115300. [CrossRef] [PubMed]
72. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [CrossRef] [PubMed]
73. Goff, N.K.; Guenther, J.F.; Roberts, J.K.; Adler, M.; Molle, M.D.; Mathews, G.; Kurouski, D. Non-Invasive and Confirmatory Differentiation of Hermaphrodite from Both Male and Female Cannabis Plants Using a Hand-Held Raman Spectrometer. *Molecules* **2022**, *27*, 4978. [CrossRef] [PubMed]
74. Mostafa, S.; Mondal, D.; Panjvani, K.; Kochian, L.; Stavness, I. Explainable Deep Learning in Plant Phenotyping. *Front. Artif. Intell.* **2023**, *6*, 1203546. [CrossRef] [PubMed]
75. Araus, J.L.; Cairns, J.E. Field High-Throughput Phenotyping: The New Crop Breeding Frontier. *Trends Plant Sci.* **2014**, *19*, 52–61. [CrossRef] [PubMed]
76. Graciano, R.P.; Peixoto, M.A.; Leach, K.A.; Suzuki, N.; Gustin, J.L.; Settles, A.M.; Armstrong, P.R.; Resende, M.F.R., Jr. Integrating Phenomic Selection Using Single-Kernel near-Infrared Spectroscopy and Genomic Selection for Corn Breeding Improvement. *Theor. Appl. Genet.* **2025**, *138*, 60. [CrossRef] [PubMed]

77. Robert, P.; Brault, C.; Rincent, R.; Segura, V. Phenomic Selection: A New and Efficient Alternative to Genomic SelectionGenomic Selection (GS). In *Genomic Prediction of Complex Traits: Methods and Protocols*; Ahmadi, N., Bartholomé, J., Eds.; Springer: New York, NY, USA, 2022; pp. 397–420. ISBN 978-1-0716-2205-6.
78. Brandariz, S.P.; González Reymúndez, A.; Lado, B.; Malosetti, M.; Garcia, A.A.F.; Quincke, M.; von Zitzewitz, J.; Castro, M.; Matus, I.; del Pozo, A.; et al. Ascertainment Bias from Imputation Methods Evaluation in Wheat. *BMC Genom.* **2016**, *17*, 773. [CrossRef] [PubMed]
79. Liu, Z.; Zhang, J.; Wang, Y.; Wang, H.; Wang, L.; Zhang, L.; Xiong, M.; He, W.; Yang, S.; Chen, Q.; et al. Development and Cross-Species Transferability of Novel Genomic-SSR Markers and Their Utility in Hybrid Identification and Trait Association Analysis in Chinese Cherry. *Horticulturae* **2022**, *8*, 222. [CrossRef]
80. Farinati, S.; Draga, S.; Betto, A.; Palumbo, F.; Vannozzi, A.; Lucchin, M.; Barcaccia, G. Current Insights and Advances into Plant Male Sterility: New Precision Breeding Technology Based on Genome Editing Applications. *Front. Plant Sci.* **2023**, *14*, 1223861. [CrossRef] [PubMed]
81. Fujita, N.; Kazama, Y.; Yamagishi, N.; Watanabe, K.; Ando, S.; Tsuji, H.; Kawano, S.; Yoshikawa, N.; Komatsu, K. Development of the VIGS System in the Dioecious Plant *Silene latifolia*. *Int. J. Mol. Sci.* **2019**, *20*, 1031. [CrossRef] [PubMed]
82. Golicz, A.A.; Batley, J.; Edwards, D. Towards Plant Pangenomics. *Plant Biotechnol. J.* **2016**, *14*, 1099–1105. [CrossRef] [PubMed]
83. Carey, S.B.; Lovell, J.T.; Jenkins, J.; Leebens-Mack, J.; Schmutz, J.; Wilson, M.A.; Harkess, A. Representing Sex Chromosomes in Genome Assemblies. *Cell Genom.* **2022**, *2*, 100132. [CrossRef] [PubMed]
84. Jayakodi, M.; Shim, H.; Mascher, M. What Are We Learning from Plant Pangenomes? *Annu. Rev. Plant Biol.* **2025**, *76*, 663–686. [CrossRef] [PubMed]
85. Farooq, M.A.; Gao, S.; Hassan, M.A.; Huang, Z.; Rasheed, A.; Hearne, S.; Prasanna, B.; Li, X.; Li, H. Artificial Intelligence in Plant Breeding. *Trends Genet.* **2024**, *40*, 891–908. [CrossRef] [PubMed]
86. Jamil, I.N.; Remali, J.; Azizan, K.A.; Nor Muhammad, N.A.; Arita, M.; Goh, H.-H.; Aizat, W.M. Systematic Multi-Omics Integration (MOI) Approach in Plant Systems Biology. *Front. Plant Sci.* **2020**, *11*, 944. [CrossRef] [PubMed]
87. Zhao, X.; Pan, S.; Liu, Z.; Han, Y.; Zhang, Q.; Wang, K. Intelligent Upgrading of Plant Breeding: Decision Support Tools in the Golden Seed Breeding Cloud Platform. *Comput. Electron. Agric.* **2022**, *194*, 106672. [CrossRef]