

Beyond the highest-yielding variety: dendrogram-based functional clustering of oil palm germplasm for multi-trait breeding, genetic diversity conservation, and climate-resilient productivity

Abstract

Oil palm breeding increasingly requires simultaneous improvement of yield, oil recovery, plant architecture, productive longevity, environmental adaptation, and genetic diversity rather than maximization of a single trait. This qualitative literature review uses as a provisional empirical anchor an unpublished dendrogram of 59 oil palm variety entries prepared by Dimas Haryo Pamungkas, Senior Researcher at IPOSS Jakarta, and separated into eight groups at a reported linkage distance of 9.71. The individual variety identities, standardized trait matrix, numerical trait values, scaling procedure, distance measure, linkage algorithm, and rationale for the 9.71 cut were not available for independent verification; accordingly, C1-C8 are treated strictly as provisional functional annotations rather than validated biological or genetic clusters. Sixty-three peer-reviewed journal sources from 2020-2026 were synthesized to examine alternative productivity pathways, trait trade-offs, genetic diversity, genomic validation, and environment-specific deployment. The literature shows that productivity can arise through different combinations of bunch number, bunch mass, biological oil-to-bunch ratio, mill extraction performance, architecture, precocity, and sustained performance, while several of these traits can trade off. The review therefore proposes a Cluster Complementarity Breeding Framework in which the dendrogram generates hypotheses for breeding pools but does not establish genetic distance, heterosis, or parental complementarity. Illustrative cross combinations, including C1 x C6 and C2 x C3, are testable examples only and require marker or pedigree/coancestry information, breeding-value and combining-ability assessment, replicated progeny trials, and evaluation across ages and environments. Phenomics, genomic prediction, GWAS, metabolomics, and machine learning are considered only insofar as they can validate, refine, or operationalize these provisional cluster hypotheses. No cluster is presented as universally superior; its breeding value depends on validated trait expression, genetic relationships, target environments, and breeding objectives.

Keywords: oil palm, *Elaeis guineensis*, dendrogram, hierarchical clustering, genetic diversity, germplasm, multi-trait breeding, genomic selection, yield components, genotype-environment interaction

Special Issue - 2026

Loso Judijanto

IPOSS Jakarta, Indonesia

Correspondence: Loso Judijanto, IPOSS Jakarta, Indonesia

Received: September 15, 2026 | **Published:** October 01, 2026

Introduction

Oil palm (*Elaeis guineensis* Jacq.) occupies an unusual position in crop improvement because it combines exceptional biological productivity with a long breeding cycle, large plant size, expensive field evaluation, strong environmental influences, and multiple economically relevant products. Improving oil yield is therefore not equivalent simply to selecting the palm with the largest fresh fruit bunch production. Oil output is the consequence of numerous interacting traits, including bunch number, average bunch weight, fruit set, mesocarp proportion, oil content, oil-to-bunch ratio, kernel characteristics, architecture, reproductive allocation, and the duration over which the palm remains economically productive. Contemporary reviews consequently argue that future oil palm improvement should integrate conventional breeding, genomics, molecular selection, phenomics and, where appropriate, genome editing rather than relying only on long cycles of phenotypic selection.¹⁻³

The need for such integration is reinforced by the economics of perennial crop breeding. Conventional evaluation requires many years before mature yield can be assessed, while the number of candidate palms that can be tested in replicated field trials is

limited. Genomic prediction has already demonstrated that selection accuracy for several oil palm yield components can be improved before candidates proceed to expensive clonal trials. Nyoma et al.,⁴ for example, reported that genomic information could improve or accelerate preselection and allow simultaneous selection for a broader set of yield components. This changes the breeding problem from one of identifying a few phenotypically impressive individuals to one of predicting which combinations of genetic backgrounds and traits will generate sustained genetic gain.

The dendrogram examined as the empirical starting point for this review was prepared by Dimas Haryo Pamungkas, Senior Researcher at IPOSS Jakarta, Indonesia, as an unpublished figure supplied for this review. It contains 59 terminal oil palm variety entries partitioned into eight groups at a reported cut height or linkage distance of 9.71 (C1 n=13; C2 n=4; C3 n=3; C4 n=10; C5 n=4; C6 n=7; C7 n=4; C8 n=14). The identities and source populations of the individual entries, the raw or standardized trait matrix, numerical trait values, scaling or normalization procedure, distance metric, linkage method, and rationale for selecting 9.71 were subject to further analysis and therefore cannot be independently assessed. The manuscript

annotations refer to relative profiles involving bunch weight, FFB/CPO output, peak duration or lifespan, production timing, kernel extraction rate (KER), extraction/recovery-related measures, bunch number, and architecture, which were inputs to the original clustering. The dendrogram structure, group sizes, and reported cut height are therefore treated as source observations, whereas C1-C8 functional descriptions are treated as provisional author interpretations for hypothesis generation.

Such an interpretation is supported by empirical oil palm genetics. Genome-wide association analyses have demonstrated that yield components are genetically complex and that different loci contribute to oil-to-bunch ratio, average bunch weight, and related traits. Babu et al.⁵ identified multiple significant loci for yield and oil-yield attributes, while Babu et al.⁶ found a negative relationship between bunch number and average bunch weight in a large African germplasm panel. These findings are important because they indicate that two palms can achieve comparable fresh fruit bunch productivity through different combinations of underlying traits.

A similar message emerges from progeny and germplasm studies. Swaray et al.⁷ demonstrated substantial variation among Dura × Pisifera progenies in fruit set, bunch number, bunch weight and oil yield, while Darkwah et al.⁸ reported exploitable agro-morphological variation among Ghanaian natural Dura accessions. More recently, the evaluation of MPOB-Angola materials documented substantial differences among families in fresh fruit bunch yield, oil yield, fruit size and oil-to-bunch ratio, illustrating why germplasm diversity should be managed as a portfolio of useful attributes rather than reduced to a single performance statistic.⁹

This distinction matters for sustainable intensification. Increasing the genetic potential of planting materials can raise output from existing agricultural land, but improved genetics alone does not determine realized yield. Evidence from Indonesian smallholder fields shows that nutrient management, pruning, weed control and harvesting frequency can account for large parts of observed yield gaps. Genetic improvement must therefore be viewed as one component of an integrated productivity system in which genotype, environment and management interact.¹⁰

The central argument of this review is consequently that oil palm breeding should move **beyond the search for a universally highest-yielding variety toward the identification and strategic combination of complementary functional trait clusters**. Phenotypic dendrograms can contribute to this shift when they are connected to molecular diversity, genomic prediction, multi-trait selection indices, environmental adaptation and long-term breeding objectives.

The review addresses four related questions. First, what does functional clustering reveal about alternative biological pathways toward oil palm productivity? Second, how should trade-offs among yield, reproductive, architectural and longevity traits influence parental selection? Third, how can phenotypic clusters be integrated with genomic and phenomic technologies? Fourth, how can cluster-based breeding simultaneously contribute to genetic gain, adaptation and diversity conservation?

Literature review

Oil Palm as a multi-trait breeding system

Oil palm differs from annual crops because the economic phenotype unfolds over a long productive life. A breeder is not selecting only a

harvest-season phenotype but a trajectory that includes establishment, first bearing, yield increase, peak production, reproductive stability, height increment and eventual economic replacement. Consequently, short-term yield superiority can differ from lifetime economic superiority. The literature on modern oil palm breeding repeatedly identifies slow vertical growth, high oil yield, stress tolerance, oil quality, reproductive performance and harvestability as simultaneous breeding objectives.^{1,3}

Fresh-fruit-bunch yield can be expressed approximately as the product of bunch number and average bunch weight, but these components are biologically interdependent rather than independently maximizable. In the African germplasm evaluated by Babu et al.,⁶ bunch number and average bunch weight were negatively correlated. Swaray et al.⁷ similarly found that different progenies excelled in different components: the progeny producing the most bunches was not necessarily the progeny with the highest average bunch weight. Genetic improvement based solely on one component may therefore shift resource allocation rather than maximize total long-run oil production.

Oil yield introduces additional layers. Mesocarp-to-fruit ratio, oil-to-mesocarp, oil-to-bunch and extraction efficiency influence how much crude palm oil is generated from a given amount of FFB. Kernel traits add another product dimension, while fatty-acid composition and micronutrient characteristics can create quality objectives distinct from tonnage. The 2026 evaluation of African Dura germplasm by Prathap et al. demonstrated wide variation in fatty-acid composition and used hierarchical clustering together with the Multi-Trait Genotype–Ideotype Distance Index to identify accessions combining nutritional-quality and agronomic characteristics. This provides direct recent evidence that oil palm germplasm can be evaluated as multidimensional ideotypes rather than along a single yield axis.

For clarity, three related but non-interchangeable quantities are distinguished throughout this review. Biological oil-to-bunch ratio (O/B) is the proportion of oil contained in the harvested bunch and is primarily a fruit-composition trait. Mill oil extraction rate (OER) is the mass of crude palm oil recovered per mass of fresh fruit bunch processed and therefore reflects biological oil content together with ripeness, transport, sterilization, processing efficiency, and mill losses. Realized oil yield is an area- and time-based output, commonly expressed as tonnes of crude palm oil per hectare per year, and depends jointly on FFB production and an explicitly defined oil-recovery measure. Because the original 59-entry trait matrix is subject to further analysis, references to C6 are restricted to an extraction/recovery-related annotation unless the exact source variable can be verified.

Genetic diversity as breeding capital

Genetic diversity is simultaneously a conservation objective and a productive asset. Intensive recurrent selection can narrow breeding populations, increase relatedness, and reduce the probability that rare but useful alleles remain available for future improvement. Studies using microsatellite and genomic markers continue to show substantial structure both within and among oil palm breeding populations. Okoye et al.,¹¹ for example, identified appreciable allelic differentiation among inter-population progenies, while Myint et al.¹² documented diversity within Senegalese germplasm using microsatellite markers. More recent conservation-oriented analyses reinforce the value of explicitly retaining allelic breadth: Gan et al.¹³ developed a reduced, molecularly informed core collection that retained the allelic diversity represented in a Nigerian-derived field genebank, whereas Zolkafli

et al.¹⁴ used genotyping-by-sequencing across African germplasm to resolve population structure and identify a compact core set that preserved the collection's allelic spectrum.

Sarimana et al.¹⁵ similarly demonstrated that Dura × Pisifera populations descended from different Pisifera origins could be differentiated using SSR markers. The importance of this evidence is not merely taxonomic. Molecular differentiation affects the design of crosses, the control of inbreeding and the possibility of exploiting complementary genetic backgrounds. Phenotypically similar palms may not be genetically close, and phenotypically different palms can sometimes share considerable pedigree. Phenotypic clusters should therefore be treated as an initial decision layer rather than as substitutes for molecular relationship information.

The value of broadening breeding populations is particularly visible in crop wild relatives. Ithnin et al.¹⁶ found substantial genetic and phenotypic diversity in wild American oil palm, *Elaeis oleifera*, including associations with fatty-acid, vegetative and yield traits. The authors emphasized the potential of wild germplasm to widen the genetic base of cultivated *E. guineensis*. Interspecific resources are especially relevant for compactness, disease resistance and oil composition, even though their deployment normally requires several generations of introgression and selection.

At the same time, genomic characterization of established breeding pools remains essential. Seyum et al.¹⁷ documented strong differentiation between key Deli and La Mé populations and showed how linkage disequilibrium and population structure influence genomic-selection design. Recent chromosome-scale assemblies for both *Elaeis guineensis* and *E. oleifera* provide a stronger genomic basis for comparative analysis and introgression,⁸ while genotyping-by-sequencing has enabled dense linkage mapping and the identification of yield-related QTLs in oil palm.¹⁹ These advances illustrate why apparent cluster complementarity should ultimately be evaluated together with genomic relationships, linkage structure, and effective population size rather than only with dendrogram branch length.

Plant architecture, compactness and economic lifespan

Architecture is sometimes treated as a secondary characteristic because it does not enter directly into FFB or oil-yield equations. In perennial palms, however, architecture has major economic consequences. Rapid vertical growth eventually increases harvesting difficulty and cost and can shorten the commercially manageable plantation life. Slow-growing or compact palms are therefore valuable even if their early annual yield is not the maximum observed.

QTL studies using interspecific backcross populations identified genomic regions associated with vegetative characteristics relevant to compactness as well as fatty-acid composition.²⁰ More recently, Samsudin et al.²¹ conducted a genome-wide association study specifically for oil palm compactness traits, reinforcing compact architecture as an explicit breeding objective. Compact palms may improve harvesting accessibility and, under appropriate agronomic conditions, allow higher planting density.

This evidence is relevant to the provisional C6 annotation, which pairs an extraction/recovery-related profile with relatively shorter architecture. Because the source matrix does not reveal whether the recovery variable was biological O/B, mill OER, CPO output, or another derived measure, C6 is not interpreted here as a verified high-OER cluster. If later source documentation confirms the metric and its

measurement basis, its breeding and economic relevance can then be evaluated against FFB production, processing conditions, and plant architecture.

Precocity and productive longevity

Precocity has clear financial significance because earlier harvesting reduces the unproductive period between planting and revenue generation. Nevertheless, an early-producing genotype is not necessarily the most productive over a full plantation cycle. Long-lived perennial systems require distinction between **early cumulative yield**, **peak annual yield** and **lifetime cumulative economic yield**.

The source annotation attached to C3 indicates a long peak period and long productive lifespan, but no longitudinal records or underlying trait matrix were available to verify this as observed productive longevity. In this review, C3 is therefore used as a testable longevity-oriented hypothesis rather than evidence of demonstrated lifespan. Productive longevity itself can arise from sustained reproductive capacity, moderate height increment, disease tolerance, physiological stability, or combinations of these traits. Breeding only for the earliest or largest initial yields could unintentionally select material whose later-life productivity declines more rapidly.

Tupaz-Vera et al.²² demonstrated that recurrent selection in Colombian oil palm populations generated genetic gain in FFB production, bunch number, and oil-to-bunch ratio while also identifying highly heritable traits such as height increment. Their findings illustrate that productivity improvement and architecture can be addressed within the same breeding program, supporting lifetime rather than purely annual optimization.

Genotype × environment interaction

Cluster membership estimated in one environment cannot automatically be interpreted as a permanent genetic category. Phenotypes reflect the combined effects of genotype, environment and their interaction. This is particularly important for reproductive traits because rainfall and water deficits can influence sex determination, inflorescence development, abortion and subsequent bunch production.

Hermanto et al.²³ used multi-season yield and rainfall data to differentiate oil palm progeny responses to water availability and identified bunch number as an especially informative parameter for studying rainfall response. Their analysis demonstrates that environmental sensitivity itself can be regarded as a selectable characteristic.

Multi-environment evidence is not necessarily expected to show strong G×E for every trait. Rachman et al.,²⁴ evaluating oil palm progenies across contrasting Indonesian environments, reported limited G×E for the bunch-component traits studied and therefore relatively stable expression for those characteristics. The lesson is not that G×E is either universally strong or universally negligible, but that stability must be estimated trait by trait and population by population.

General crop-breeding research provides a broad statistical framework for this requirement. Stability indices, multi-environment genomic prediction and sparse-testing designs have become central to identifying genotypes that combine performance with stability.^{25,26} Genotype–environment association studies can also connect environmental gradients with genomic variation underlying adaptation.²⁷

Methods

Review design

This study applies a **qualitative literature review (QLR)** with narrative and thematic synthesis. It is deliberately distinct from a systematic literature review. The objective is therefore not to claim exhaustive retrieval of every publication or to estimate pooled quantitative effect sizes. Instead, recent studies are used to interpret the biological and breeding significance of the supplied 59-variety dendrogram and to construct a conceptual framework linking varietal clustering with contemporary breeding technologies.

The temporal scope was 2020-2026. Literature identification was purposive and iterative rather than exhaustive, using searches across Scopus, Web of Science, PubMed, ScienceDirect, SpringerLink, Wiley Online Library, and Google Scholar. Search combinations paired “oil palm” or “*Elaeis guineensis*” with terms including “genetic diversity”, “germplasm”, “cluster analysis”, “yield components”, “genomic selection”, “GWAS”, “QTL”, “phenotyping”, “genotype x environment”, “machine learning”, and “multi-omics”. Titles and abstracts were first screened for direct relevance to the review questions; potentially relevant papers were then assessed at full-text level for population and trait relevance, methodological transparency, and contribution to one or more analytical themes. Reference chaining and additional targeted searches were used when new subthemes emerged. Because the search was not prospectively logged as a systematic review, the number of initially retrieved or screened records cannot be reconstructed reliably and no PRISMA-style flow count is claimed.

Eligible sources were peer-reviewed journal articles published from 2020 through 2026 that directly addressed oil palm breeding, germplasm, yield, architecture, adaptation, genomic tools, or deployment, together with methodologically relevant crop-breeding studies when oil-palm-specific evidence was limited. The final narrative synthesis retained 63 peer-reviewed sources cited in this manuscript. Evidence was prioritized in the following order: direct empirical oil palm studies, oil-palm-specific reviews, and then transferable methodological studies from other crops. Conflicting findings were not pooled into a single effect; instead, they were

compared by trait definition, germplasm or population, environment, palm age, and study design. Where oil palm studies differed, both findings were retained and interpreted as context-dependent, as illustrated by the contrasting strength of genotype x environment effects reported for different traits and populations. Conference abstracts, non-peer-reviewed opinion pieces, duplicate reports, and sources without sufficient methodological detail were not used as substantive evidence.

Thematic coding and evidence organization

The evidence was organized around eight analytical themes: (1) diversity and germplasm structure; (2) alternative pathways to yield; (3) trade-offs among yield components; (4) oil and kernel recovery; (5) architecture, precocity, and longevity; (6) molecular and genomic validation; (7) genotype x environment interaction; and (8) cluster-informed multi-trait breeding. These themes were treated as interpretive coding categories rather than as mutually exclusive bins, allowing individual studies to inform more than one theme when their results spanned multiple breeding objectives.

Throughout the synthesis, three evidence layers are kept separate: (i) observations directly available from the supplied dendrogram, such as the tree structure, group sizes, and reported cut height; (ii) findings reported in peer-reviewed studies; and (iii) author-generated interpretations or proposals, including the functional C1-C8 labels, illustrative cross combinations, and the conceptual pipelines in Figures 2-4. This separation is used to prevent interpretive proposals from being presented as empirical properties of the 59 entries.

The supplied dendrogram was treated as an empirical anchor rather than an independent experimental dataset analyzed *de novo* in this review. The figure reports 59 terminal entries, eight groups, a cut height of 9.71, and the cluster sizes summarized in Table 1. However, individual variety identities, original population provenance, the complete trait list used to calculate distances, raw or standardized trait values, scaling, the distance metric, linkage algorithm, the rationale for the 9.71 cut, and clustering code were unavailable. No attempt was therefore made to reconstruct membership or infer unreported cluster means. Phenotypic branch separation is not interpreted as genomic distance, and C1-C8 are treated strictly as provisional annotations pending to the original data and methods.

Table 1 Evidence map for the provisional C1-C8 annotations derived from the supplied 59-entry dendrogram

Cluster (n)	Trait direction used in the manuscript	Evidence status	Interpretive constraint
C1 (13)	Relatively large bunch weight; strong harvest-age performance.	Functional annotation	Provisional bunch-mass hypothesis
C2 (4)	High FFB/CPO output; early and extended peak.	Functional annotation	Provisional high-output hypothesis
C3 (3)	Long peak period; long productive lifespan.	Functional annotation	Longevity-oriented hypothesis
C4 (10)	Early production; relatively high palm-product output.	Functional annotation	Provisional precocity/productivity hypothesis.
C5 (4)	Distinct reproductive profile; relatively low KER.	Functional annotation	Provisional reproductive/kernel-allocation hypothesis.
C6 (7)	High extraction/recovery-related annotation and shorter architecture.	Functional annotation.	Provisional recovery/compactness hypothesis
C7 (4)	Many bunches with relatively lower bunch weight and recovery-related performance.	Functional annotation	Provisional bunch-number hypothesis
C8 (14)	Relatively high KER with intermediate productivity.	Functional annotation	Provisional intermediate/kernel-oriented hypothesis

Note: only the dendrogram grouping, branch structure, reported cut height, and cluster sizes are directly observable from the dendrogram figure. The functional descriptions are provisional interpretations; individual variety identities and raw cluster-defining trait values were subject to further independent verification.

Analytical logic

The interpretive process proceeded from source observation to testable breeding hypothesis. First, only the information visible or stated in the supplied dendrogram was recorded. Second, the provisional functional annotations were compared with recent literature to determine whether the implicated traits have demonstrated heritable variation, genomic associations, or breeding relevance. Third, possible trade-offs and complementarities were identified from published evidence rather than inferred from branch distance. Fourth, the review considered how the provisional groups could be tested using genomic, pedigree, phenomic, longitudinal, and environmental information before any crossing decision is made.

This approach is consistent with the broader shift toward multivariate breeding. Multi-trait genomic-selection frameworks attempt to improve simultaneous response across correlated traits rather than optimize each trait separately.²⁸ MGIDI similarly identifies genotypes by their multivariate distance from a predefined ideotype rather than using a single measurement.²⁹

Results: thematic findings

Theme I — varietal diversity is functionally structured

The first synthesis finding is that oil palm diversity is not random noise around a universal ideal. It is frequently structured into recurring functional combinations. Some materials produce yield through larger bunches, others through greater bunch number, others through superior oil content, and still others offer architectural or quality characteristics that may be economically valuable despite moderate FFB.

Given the source limitations summarized in Table 1, the eight groups are discussed only as provisional functional annotations. For hypothesis generation, C1 is treated as bunch-mass-oriented; C2 as comparatively high-output; C3 as longevity-oriented; C4 as early-return-oriented; C5 as a distinctive reproductive/kernel-allocation profile; C6 as extraction/recovery-related plus compactness; C7 as bunch-number-oriented; and C8 as intermediate or kernel-oriented. These labels organize the narrative synthesis but are not presented as independently estimated cluster means or stable genetic classes.

Similar multivariate differentiation has been observed elsewhere. Darkwah et al.⁸ found meaningful variation across 18 agromorphological traits in Ghanaian Dura accessions. Swaray et al.⁷ used UPGMA and principal-component analysis to distinguish progeny groups with different combinations of fruit set and oil yield. Prathap et al.³⁰ likewise found clear multivariate differentiation among African germplasm for oil composition, quality and agronomic traits.

The functional interpretation is summarized in Figure 1. The figure does not imply a linear ranking among clusters. Instead, it represents each cluster as a different route through which germplasm may contribute to a breeding program.

Figure 1 reframes the supplied dendrogram as a set of testable breeding hypotheses rather than a taxonomy of verified functional types. For example, C1 may warrant testing for bunch-mass contributions and C7 for bunch-number contributions, whereas the C6 annotation warrants verification of both architecture and the exact recovery variable before any claim of high OER is made. Similarly, C3 should be evaluated with longitudinal records before being treated as a longevity donor. The breeding question is therefore which provisional functions merit validation and combination.

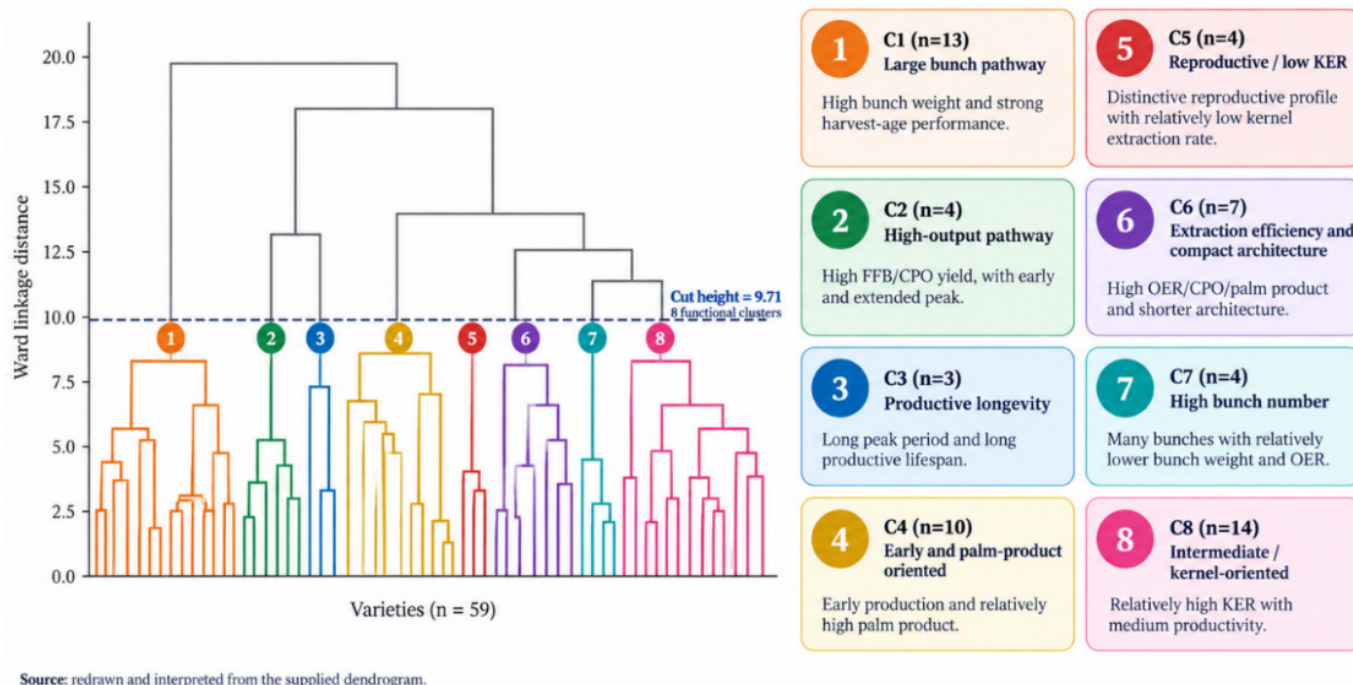


Figure 1 Provisional functional interpretation of the supplied 59-entry oil palm dendrogram. The tree structure, eight groups, cluster sizes, and reported cut height of 9.71 are source observations; the colored functional labels are author-generated annotations used for hypothesis generation. Individual variety identities, raw trait values, scaling, distance metric, linkage method, and the cut-height rationale were subject to further analysis, so the labels are not independently validated cluster means and do not constitute a superiority ranking.

Theme 2 — high yield has multiple biological pathways

Yield-component studies consistently show that FFB can be assembled through different trait combinations. Babu et al.⁶ identified a negative correlation between bunch number and average bunch weight, illustrating a biological trade-off between reproductive frequency and individual bunch mass. This is directly relevant to the contrast between C1 and C7 in the supplied dendrogram. C1 appears to emphasize bunch size, whereas C7 emphasizes bunch number. Neither pathway can be judged without considering total annual FFB, oil fraction, harvest cost, and stability.

A second pathway may operate through oil recovery, but biological O/B, mill OER, and realized oil yield must remain separate. A genotype with moderate FFB may still achieve high area-based oil output when a verified biological oil fraction and/or mill extraction performance is favorable; realized oil yield can be inferred only when FFB production and the relevant recovery measure are both defined. The C6 source annotation includes OER/CPO/palm-product terminology, but the unavailable matrix prevents determination of which variable was actually measured. C6 is therefore discussed as a provisional extraction/recovery-related and shorter-architecture profile, not as a confirmed high-OER group.

Babu et al.⁵ found multiple loci associated with oil-to-bunch ratio and average bunch weight, supporting the view that these traits can be genetically differentiated. Norziha et al.⁹ similarly identified oil palm germplasm families that differed substantially in FFB, oil yield, and oil-to-bunch ratio. In addition, Bhagya et al.³¹ reported marker-trait associations for several oil-yield components in diverse oil palm germplasm, further demonstrating that component traits can be genetically dissected and recombined. Together, these studies provide biological justification for treating cluster performance as multidimensional rather than as a single yield ranking.

Theme 3 — trait trade-offs are central, not peripheral

The third finding is that trade-offs are integral to oil palm breeding. The clearest example is bunch number versus bunch mass, but other trade-offs are plausible between vegetative growth and reproductive allocation, early yield and sustained yield, kernel allocation and mesocarp oil, or high-density architecture and individual-palm canopy development.

This makes independent single-trait ranking problematic. Multi-trait genomic-selection research emphasizes that breeders typically face simultaneous objectives, some of which are antagonistically correlated. Moeinizada et al.²⁸ showed conceptually how multi-objective genomic selection can balance several traits rather than maximize one at the expense of others. MGIDI provides another framework by selecting genotypes closest to a multidimensional ideotype.²⁹

Recent crop studies reinforce the same principle. Multi-trait genomic prediction can improve prediction accuracy by using information from correlated traits, as demonstrated in wheat by Shahi et al.³² and in rice by Muvunyi et al.³³ Although these crops have very different biology from oil palm, the statistical principle is directly transferable: correlated secondary traits can improve prediction of expensive or difficult primary traits.

Theme 4 — productive longevity is a distinct breeding objective

C3 is useful as a hypothesis because its annotation points to productive persistence rather than only annual output; however, the supplied material does not contain longitudinal records that

demonstrate observed longevity. For a perennial crop, verified longevity would affect cumulative oil yield, replanting intervals, labor requirements, plantation renewal costs, and the period during which establishment capital continues generating revenue. The C3 label should therefore be treated as a proposed long-term breeding objective to be tested, not as an established phenotype of the three entries.

Productive longevity should therefore be represented by longitudinal traits rather than a single annual observation. Moreira et al.³⁴ emphasize the value of combining longitudinal phenotyping with statistical genomics because dynamic traits cannot always be represented adequately by endpoint measurements. This is highly relevant to oil palm, where the economically meaningful phenotype includes a yield curve extending across many years.

Longitudinal data can also distinguish environmentally driven fluctuation from genetic decline. Hermanto et al.²³ demonstrated how multi-season oil palm yield records can be decomposed to identify rainfall-sensitive reproductive stages and progeny-specific responses. A future breeding interpretation of C3 should consequently include cumulative yield, yield persistence, height increment and stress response, not merely the number of years classified as “productive.”

Theme 5 — compactness creates indirect but substantial value

Compactness offers a case in which the largest breeding benefit may occur indirectly. A shorter palm can remain harvestable for longer, reduce harvesting difficulty, improve worker productivity, and potentially permit higher planting densities. QTL analysis in interspecific backcrosses and recent GWAS both show that vegetative compactness has a tractable genetic component.^{20,21}

These characteristics make the C6 hypothesis worth testing, but donor value cannot be assigned until the source recovery metric is verified. If C6 entries are subsequently shown to combine compact architecture with a clearly measured favorable O/B, mill OER, or area-based oil yield, their economic value could be evaluated over the plantation life cycle. This illustrates why trait definitions and measurements must be established before cluster annotations are converted into breeding or economic indices.

Theme 6 — phenotypic clustering requires molecular validation

Phenotypic clustering is useful because it simplifies multidimensional information, but it does not establish underlying genetic distance. Environmental effects, age, management, and measurement scales can alter phenotypic similarity. Molecular characterization is therefore a necessary second layer. SSR-based studies by Myint et al.¹² and Sarimana et al.¹⁵ demonstrate how molecular markers distinguish oil palm population structure, while Seyum et al.¹⁷ show how genomic properties such as linkage disequilibrium influence genomic-selection performance. Genomic diversity analysis of *E. oleifera* further shows that genetically informative wild material can contain useful yield, quality and vegetative variation that may not be evident from conventional commercial pedigrees alone.¹⁶

Figure 2 is an author-proposed conceptual pipeline showing how a provisional phenotypic grouping could be converted into a testable breeding decision. The supplied dendrogram generates hypotheses about functional complementarity; marker, pedigree, phenotypic, longitudinal, and environmental evidence must determine whether those hypotheses are retained, revised, or rejected before crosses are made.

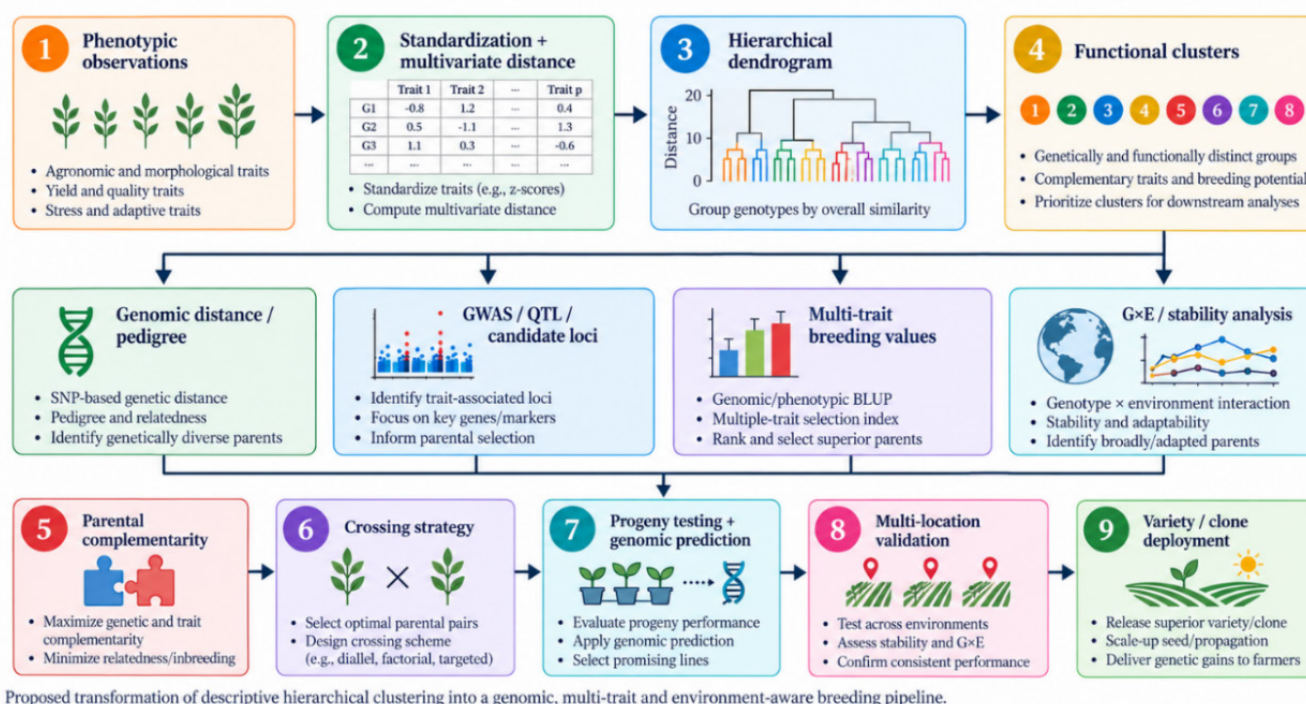


Figure 2 Author-proposed conceptual pipeline for transforming a descriptive dendrogram into a genomic, multi-trait, and environment-aware breeding decision process. The workflow is a synthesis proposed in this review, not a procedure demonstrated with the 59-entry dataset.

The critical step in Figure 2 is the transition from provisional cluster membership to validated parental complementarity. Phenotypic separation alone does not establish genetic distance, complementarity, or heterosis. Candidate parents should be evaluated with marker data and pedigree or coancestry, predicted breeding values where appropriate, combining-ability evidence, and adaptation to the target environment. Tchounke et al.³⁵ show that mate-selection approaches can explicitly balance expected genetic gain with control of future inbreeding, which supports using genomic relatedness as an additional decision layer rather than treating dendrogram branch length as a mating rule.

Discussion and analysis

From eight clusters to a functional breeding portfolio

The supplied dendrogram visually partitions the 59 entries into eight groups, but the unavailable matrix and clustering metadata prevent those groups from being assumed to be stable biological or genetic resource pools. Their value in this review is narrower: they provide a structured set of provisional hypotheses that can be matched to breeding objectives and then tested with the original phenotypic data, genomic relationships, and longitudinal performance.

Subject to that qualification, C1 is provisionally interpreted as bunch-mass-oriented; C2 as comparatively high-output; C3 as a longevity-oriented hypothesis; C4 as precocity-oriented; C5 as a distinctive reproductive/kernel-allocation profile; C6 as extraction/recovery-related plus compactness; C7 as bunch-number-oriented; and C8 as intermediate or kernel-oriented. These descriptions guide questions for validation rather than assert verified donor functions.

This interpretation remains provisional not only because the dendrogram is phenotype-based, but also because the original trait matrix, identities, scaling, and clustering method were unavailable.

A useful breeding framework can nevertheless be built if these labels are treated as prompts for measurement and validation rather than as evidence of allelic content or genetic superiority.

Within-cluster and between-cluster selection

Two complementary strategies can be tested once original identities and trait measurements are recovered. Within-group selection could identify candidates that consistently express a desired verified profile, such as compact architecture together with a clearly defined recovery trait among C6 entries. Between-group crossing could then be considered only after phenotypic differences are confirmed and genetic relationships are quantified.

Between-group crossing is therefore framed here as hypothesis testing rather than as a consequence of phenotypic distance. The objective would be to recombine verified complementary traits while controlling relatedness and maintaining sufficient diversity. Molecular studies showing differentiation among oil palm populations provide the tools needed to evaluate whether two phenotypically distinct parents are also genetically appropriate for crossing.^{11,15,17}

Illustrative test cases include C1 × C6, which asks whether a verified bunch-mass donor can be combined with a verified recovery/compactness donor; C2 × C3, which asks whether high current output can be combined with sustained performance; C4 × C3, which tests early bearing together with long-term productivity; and C7 × C6, which tests frequent bunch production together with a verified recovery trait and manageable architecture. These examples are author-generated experimental questions, not predictions that the crosses will be superior or heterotic.

Each example should pass a staged validation sequence before advancement: first recover and audit the original variety identities and trait matrix; next quantify genomic relationships with marker data and, where available, pedigree or coancestry; then estimate parental

breeding values and conduct combining-ability tests; subsequently evaluate replicated progeny across relevant sites; and finally assess performance repeatedly across age phases and production environments. Only after these stages would it be defensible to infer useful complementarity, heterosis, stability, or lifetime advantage. Transgressive segregation, epistasis, maternal effects, linkage

drag, and genotype x environment interaction remain alternative explanations for progeny outcomes.

To make the distinction explicit, Figure 3 presents a cluster-complementarity matrix. It illustrates the type of breeding logic that can be developed from the dendrogram without claiming that any specific cross is already proven.

Donor combination	Primary trait complementarity	Proposed breeding objective
C1 × C6	Large bunch × OER/compactness	High oil output with harvestable architecture
C2 × C3	High output × longevity	High lifetime cumulative productivity
C4 × C3	Precocity × longevity	Early return plus sustained production
C7 × C6	High bunch number × extraction efficiency	Frequent production with better oil recovery
C1 × C7	Bunch mass × bunch number	Recombine alternative FFB pathways
C8 × C6	Kernel/balanced profile × high OER	Diversified oil and kernel value
C2 × C6	High output × compactness	Intensification-oriented ideotype
C3 × C6	Longevity × compactness	Extended economic harvesting life

Hypothesis-generating examples of functional cluster complementarity. Proposed crosses require genomic-distance, breeding-value, combining-ability and progeny validation.

Figure 3 Author-generated, hypothesis-generating examples of functional cluster complementarity. C1 × C6, C2 × C3, and the other illustrated combinations are experimental test cases only; phenotypic cluster separation does not establish genetic complementarity or heterosis. Advancement requires source-data verification, marker/pedigree or coancestry assessment, breeding-value and combining-ability evaluation, replicated progeny trials, and performance assessment across ages and environments.

The matrix demonstrates why cluster analysis is most valuable when it leads to explicit breeding hypotheses. Rather than retaining eight descriptive groups in a report, breeders can use the groups to prioritize which crosses deserve genomic evaluation and limited field resources. This is particularly important in oil palm because each breeding cycle is expensive and slow.

Toward a multi-trait oil-palm ideotype

An ideotype for next-generation oil palm should not be defined merely as a palm producing the maximum annual FFB. A more defensible ideotype would combine high oil yield, sufficient bunch number, manageable bunch weight, high oil-to-bunch ratio, suitable kernel composition, slow height increment, harvestable architecture, early bearing, long productive lifespan, yield stability and resistance or tolerance to major stresses.

Yue, Ye and Lee¹ similarly propose an “ideal palm” incorporating high oil yield, slow height growth, resilience, harvesting suitability, efficient nutrient use, improved oil quality and longer utilization. Contemporary oil palm breeding therefore already contains the conceptual foundations for a multi-trait ideotype.

The challenge is operational. Economic weights change across production systems. A plantation with severe labor scarcity may assign greater value to compactness and harvestability. A processor constrained by mill throughput may prioritize higher oil recovery per tonne of FFB. A region with increasing water deficit may emphasize yield stability. Smallholders concerned with early cash flow may place greater value on precocity. Accordingly, the “ideal” genotype

should be target-population-of-environments and production-system specific.

Genomic selection as the bridge from clustering to prediction

Genomic prediction is directly relevant to the cluster framework because oil-palm-specific studies have already demonstrated genomic selection for clonal preselection and parental decision-making. Nyouma et al.⁴ showed improved clonal selection in oil palm hybrids, Tchoune et al.³⁵ demonstrated mate-selection strategies that balance genetic gain and inbreeding, and Garzon-Martinez et al. (2022) showed multi-trait genomic prediction in OXG and backcross populations. These are demonstrated oil palm applications, but none validates the C1-C8 labels; rather, they provide tools for testing whether phenotypically annotated groups contain useful breeding values and relationships.

Operationally, a provisional group label should enter a genomic model only as contextual phenotype information, not as a final selection rule. For example, if C6 compactness and its recovery metric are verified, genomic estimated breeding values could help identify stronger candidates within that group, while marker relationships or pedigree could test relatedness to candidate C1 parents. Selection would then rest on measured phenotype, predicted breeding value, and quantified relatedness rather than on dendrogram distance alone.

General genomic-selection research further shows that prediction accuracy depends on training-population composition, marker density, linkage disequilibrium, trait heritability, and the relationship

between training and target populations.³⁶ These cross-crop principles are relevant to model design, but approaches such as speed breeding are not directly transferable to a long-lived palm and are not used here as evidence for the cluster framework.³⁷

Phenotyping must catch up with genotyping

Within this framework, phenomics is relevant only insofar as it can test specific provisional attributes that are currently unverified, such as C3 persistence or C6 architecture. General crop studies show that high-throughput phenotyping can repeatedly measure structure, physiology, and environmental response.^{38–41} These sources establish methodological capability, not validation of the eight oil palm groups.

For oil palm, a targeted application would be repeated measurement of height increment, crown structure, bunch number, canopy status, and stress response so that the provisional labels can be tested across age. LiDAR and remote sensing are promising for three-dimensional architecture and repeated canopy assessment, but much of the supporting phenomics literature remains cross-crop rather than specific to the present oil palm material.^{42–45}

The practical contribution of phenomics to this review is therefore a reproducible trait matrix: standardized acquisition, processing, metadata, and longitudinal measurements that can be linked back to each variety entry. Data-workflow principles described by Thompson et al.⁴⁶ and Kim⁴⁷ are useful for that purpose, but they should be prospectively validated in oil palm before operational adoption.

Machine learning and digital breeding

Machine learning can contribute by predicting measured traits and testing whether multivariate groupings remain stable when genotype, phenotype, age, and environment are analyzed jointly. An oil-palm-specific example exists: Shin et al.⁴⁸ used marker information and machine-learning algorithms to predict phenotypic values in interspecific hybrid populations. This supports prediction as a validation tool, but it does not justify inferring genetic complementarity or heterosis from the current C1–C8 labels.

Deep learning and ensemble methods are discussed more cautiously because their supporting evidence is largely general plant-genomic literature. Their performance depends on sample size, genetic architecture, and validation design.^{49,50} With only 59 source entries and no accessible matrix, complex models would be inappropriate at present; simpler, externally validated models may be preferable until a larger oil palm dataset is assembled.

The useful longer-term direction is a digital breeding record in which each entry has linked phenotype, pedigree, marker, environmental, and longitudinal data. Jeon et al.⁵¹ describe this broader digital-breeding concept. In such a system, C1–C8 could be retained only as historical or interpretable labels and revised whenever the accumulating data no longer support them.

Metabolomics and multi-omics

Metabolomics has a narrower role in the cluster framework: it can test whether provisionally distinct groups also differ in biochemical signatures relevant to oil quality or stress response. Nugroho et al.⁵² specifically discuss metabolomics-assisted breeding in oil palm, making this a directly relevant oil-palm application; it is not, however, evidence that the current cluster labels are biochemically distinct.

Cross-crop metabolomics research supports integration of biochemical phenotype with genomic information.⁵³ For this review, such evidence is used only to motivate future validation of specific

cluster hypotheses rather than to infer mechanisms from the present dendrogram.

Multi-omics could be added after robust phenotype and genomic differences are established, for example to investigate mechanisms underlying verified longevity or compactness traits. Integrated omics infrastructures described in crop research can inform study design,^{54,55} but their transfer to oil palm and to the 59 source entries remains prospective.

Accordingly, a verified C3 longitudinal phenotype could later be examined for genomic or metabolic signatures, and a verified C6 compactness phenotype could be linked to architectural QTLs or three-dimensional phenomics. These would be follow-up tests of measured traits, not mechanisms assigned in advance from the dendrogram label.

Somatic embryogenesis and deployment of elite genotypes

Somatic embryogenesis is relevant only at the downstream deployment stage, after elite individuals have been validated by phenotype and genetics. It does not provide evidence for the existence or usefulness of the current clusters. Oil palm cannot be propagated clonally by conventional cuttings, so tissue culture can be valuable for multiplying selected material once its performance has been established.

Astari et al.⁵⁶ demonstrated substantial genotypic effects on oil palm somatic-embryogenesis success, supporting its role in multiplication of selected breeding material. Within the present framework, this technology belongs after parental/progeny validation and field stability testing rather than within the evidence used to define C1–C8.

Genome editing: a longer-term extension

Genome editing is a longer-term extension and is not evidence for the present cluster framework. Much of the demonstrated precision-editing evidence cited here comes from annual crops.⁵⁷ For oil palm, the long juvenile period and transformation constraints mean that editing should be considered only after causal loci and regulatory mechanisms are established in oil-palm-specific material.

The sequence is therefore discovery and validation first, intervention second: recover the original phenotypes; verify cluster stability; map useful loci with GWAS or QTL approaches; test genomic prediction and mating decisions; and only then consider whether a validated causal target is better addressed by conventional recombination or, where feasible, editing. This keeps genome editing as a prospective tool rather than a general technology overview disconnected from the cluster question.

Climate-resilient cluster deployment

Climate resilience adds another reason to retain rather than eliminate diversity. A genotype performing exceptionally in one rainfall or soil environment may not be the safest choice across a geographically heterogeneous plantation system. Phenotypic plasticity means that a single genotype can express different phenotypes under different conditions, and breeding programs must distinguish useful adaptation from undesirable instability.⁵⁸

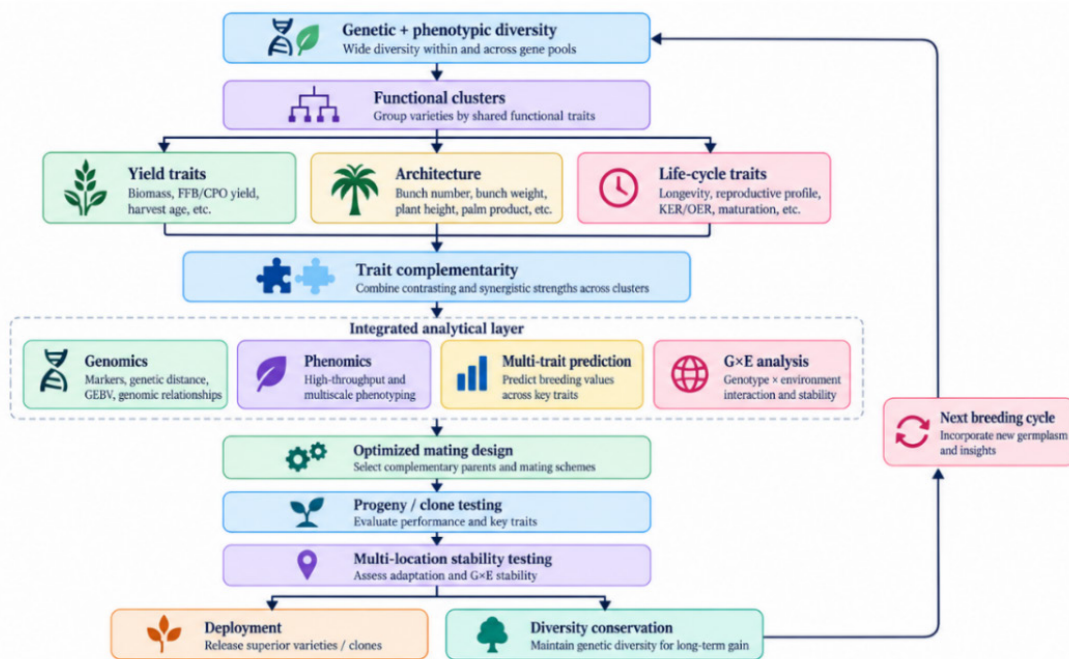
Genomic selection for climate resilience increasingly incorporates multi-environment data and stress-related traits. Budhlakoti et al.⁵⁹ review how genomic-selection systems can accelerate breeding for abiotic and biotic stress resistance. Multi-environment genomic prediction and genotype-environment association models provide a

route for identifying both broadly adapted and specifically adapted material.

For oil palm, the implication is that cluster labels should eventually acquire an environmental dimension. C2 might be subdivided into high-output material that is broadly stable and high-output material that performs especially well under a particular moisture regime. C3 longevity may differ between areas with high disease pressure and

relatively benign environments. C6 compactness may be especially attractive in high-density systems but require evaluation for canopy competition under different radiation regimes.

Figure 4 summarizes the proposed full framework. It adds environmental validation and diversity conservation to the conventional sequence from selection to release.



Proposed Cluster Complementarity Breeding Framework integrating phenotypic clustering, genomics, phenomics, multi-trait prediction, G×E analysis, optimized mating and genetic-diversity conservation.

Figure 4 Author-proposed Cluster Complementarity Breeding Framework integrating provisional phenotypic grouping with genomics, phenomics, multi-trait prediction, G×E analysis, optimized mating, progeny validation, and genetic-diversity conservation. The framework is a conceptual synthesis proposed in this review and has not been validated with the 59-entry dataset.

The feedback loop at the bottom of Figure 4 is critical. Elite commercial selection should not cause the remaining germplasm to be discarded. Rare alleles, unusual trait combinations, and wild or semi-wild accessions may become valuable when climatic, disease, labor, or quality objectives change. Genetic conservation and genetic gain are therefore complementary elements of long-term breeding rather than competing goals.

Implications for sustainable intensification

Improved planting materials can contribute to sustainable intensification by raising output without requiring proportional land expansion. However, sustainable intensification cannot be reduced to genetics alone. Monzon et al.¹⁰ demonstrated that large yield gaps in Indonesian smallholder systems are strongly associated with agronomic management, while Cock, Ayala-Diaz and Romero²² emphasized the value of distributing genetically improved material that is well matched to local production conditions. Superior genetics deployed under poor nutrient management, irregular harvesting, or unsuitable environments may therefore never express their full potential.

The interaction among genetics, environment, and management is presented here only as a conceptual heuristic, not as a fitted production function. Equation (1) is a schematic multiplicative

illustration intended to emphasize joint constraint by genetic potential, environmental suitability, and agronomic realization; its coefficients, dimensional form, and multiplicative structure have not been estimated or validated from data and it should not be used to calculate or predict productivity.

$$P_r = G_p \times E_p \times A_r \quad (1)$$

In Equation (1), P_r denotes realized productivity, G_p genetic potential, E_p environmental suitability, and A_r agronomic realization. The symbols represent interacting domains rather than independently measured factors in a calibrated model. Cluster-informed breeding can alter genetic potential and its fit to target environments, whereas nutrient management, harvesting, and plantation management affect how much of that potential is realized. This qualitative production-system perspective is consistent with evidence that improved planting material performs best when matched to local conditions and effective agronomic management.^{60,10}

This also explains why high FFB should not be the sole sustainability metric. A genotype that produces somewhat less FFB could still generate greater usable oil per hectare over a plantation cycle if it combines a verified favorable recovery measure, slower height increment, longer harvestable life, and stable output. Which recovery measure matters must be stated explicitly: biological O/B

describes bunch composition, mill OER describes recovered CPO per unit FFB, and realized oil yield describes output per area and time. Life-cycle productivity should therefore be evaluated with clearly defined component metrics rather than with interchangeable use of these terms.

Future research agenda

Five research priorities emerge from this synthesis. First, the 59 varieties represented in the supplied dendrogram should be characterized using molecular markers or genome-wide SNP information so that phenotypic and genomic distances can be compared directly. Second, the dendrogram should be reconstructed by environment and age class. Cluster stability over time would indicate whether the observed groupings represent persistent functional differences or temporal developmental states.

Third, cumulative oil productivity should be estimated over long time horizons. C3 in particular provides an opportunity to test whether productive longevity delivers greater discounted lifetime output than high early yield.

Fourth, cluster complementarity should be tested experimentally through a staged design. The original identities and phenotypic matrix should first be recovered and audited; marker data and pedigree or coancestry should then quantify genetic relationships; parental breeding values and combining ability should be estimated; and a limited set of inter-group and within-group crosses should be evaluated in replicated progeny trials across sites, years, and age phases. Heterosis, segregation, transgressive performance, and stability should be inferred from those progeny data rather than from phenotypic branch distance.

Fifth, multi-trait prediction should be integrated with longitudinal phenotyping. Advances in crop phenomics demonstrate that repeated remote-sensing measurements can help characterize dynamic phenotypes that traditional end-point measurements miss.^{34,38} Such an approach is especially appropriate for long-lived perennial crops.

Limitations of the review

This review is qualitative and interpretive; it does not claim exhaustive coverage of the literature or quantitative estimation of pooled effects. The dendrogram of 59 entries is used as an empirical anchor, but the individual variety identities, source-population details, original trait matrix, numerical cluster summaries, scaling, distance metric, linkage method, rationale for the 9.71 cut, complete linkage calculations, and clustering code were not available for independent reanalysis. Consequently, the C1–C8 labels, trait directions, and inter-group crossing examples are provisional breeding propositions that require validation against the original phenotypic data, genomic relationships, pedigree/coancestry, combining ability, replicated progeny performance, and longitudinal multi-environment observations.

A second limitation concerns transferability across crops and environments. Direct oil-palm evidence supports several components discussed here, including genomic selection, marker-based population analysis, compactness genetics, machine-learning prediction, metabolomics-assisted breeding, and somatic embryogenesis. By contrast, parts of the phenomics, deep-learning, integrated multi-omics, and genome-editing discussion rely substantially on broader crop literature because oil-palm-specific evidence remains less abundant. These approaches are therefore presented as candidate

validation or deployment tools, not as demonstrated solutions for the 59 entries. Prospective oil palm studies in target populations and environments are required before operational adoption.^{61–63}

Conclusion and recommendations

This review shows that a dendrogram can be useful for organizing breeding questions, but the present 59-entry figure does not provide enough underlying information to validate functional cluster identities. The eight groups can only be provisionally annotated as bunch-mass-oriented, comparatively high-output, longevity-oriented, precocity-oriented, reproductive/kernel-allocation, extraction/recovery plus compactness, bunch-number-oriented, and intermediate/kernel-oriented profiles. These descriptions are hypotheses for verification, not empirical cluster means or genetic classes.

The literature synthesis nevertheless supports the broader biological point that high productivity can arise from alternative combinations of bunch number, bunch mass, biological oil fraction, mill recovery, architecture, and sustained performance. Commercially important traits such as compactness, stability, and long-term productivity are not adequately represented by short-term FFB ranking alone, but each component must be defined and measured explicitly.

The proposed Cluster Complementarity Breeding Framework is therefore a hypothesis-testing sequence rather than a rule for choosing parents. Phenotypic grouping can identify questions worth investigating, but genomic relationships, pedigree or coancestry, breeding values, multi-trait indices, GxE responses, and combining ability should determine whether candidate parents proceed to crosses. Illustrative combinations such as C1 x C6 and C2 x C3 are testable examples only and should not be interpreted as predicted superior or heterotic crosses.

A defensible validation pathway is to recover the original identities and trait matrix, quantify genetic relatedness with markers and pedigree/coancestry, estimate parental breeding values, test combining ability, conduct replicated progeny trials, and follow performance across ages and environments. Only after that sequence should functional complementarity, heterosis, or lifetime advantage be claimed.

A priority research direction is therefore to combine the supplied dendrogram with its original phenotypic matrix, SNP genotyping, pedigree information, longitudinal yield records, plant-architecture phenomics, and multi-environment testing. Such integration would show whether C1–C8 correspond to stable genetic-functional populations, environmentally contingent phenotypes, age-dependent groupings, or combinations of these possibilities.

Breeding programs should also preserve representative material from strategically important clusters. Eliminating moderate-yield germplasm solely because it is not currently elite risks losing alleles for compactness, oil quality, stress adaptation or longevity that may become essential under future production conditions.

The broader implication is that genetic diversity should be viewed as a productive portfolio. The objective is not merely to conserve diversity, nor simply to maximize current yield, but to continually convert complementary diversity into genetic gain while maintaining enough variation for subsequent breeding cycles. For a perennial crop facing changing climate, labor constraints and increasing pressure to raise production from existing land, that balance is central to long-term oil palm productivity.

Acknowledgements

None.

Conflicts of interest

The author declares there is no conflict of interest.

References

1. Yue GH, Ye BQ, Lee M. Molecular approaches for improving oil palm for oil. *Mol Breed*. 2021;41:22.
2. Babu BK, Mathur RK, Anitha P, et al. Phenomics, genomics of oil palm (*Elaeis guineensis* Jacq.): way forward for making sustainable and high yielding quality oil palm. *Physiol Mol Biol Plants*. 2021a;27:587–604.
3. Martin JJJ, Yarra R, Wei L, et al. Oil palm breeding in the modern era: Challenges and opportunities. *Plants*. 2022;11:1395.
4. Nyouma A, Bell JM, Jacob F, et al. Genomic predictions improve clonal selection in oil palm (*Elaeis guineensis* Jacq.) hybrids. *Plant Sci*. 2020;299:110547.
5. Babu BK, Mathur RK, Ravichandran G, et al. Genome wide association study and identification of candidate genes for yield and oil yield related traits in oil palm using SNPs by genotyping-based sequencing. *Genomics*. 2020;112(1):1011–1020.
6. Babu BK, Mathur RK, Venu MVB, et al. Genome-wide association study of major QTLs for bunch and oil yield related traits in *Elaeis guineensis* L. *Plant Sci*. 2021b;305:110810.
7. Swaray S, Amiruddin MD, Rafi MY, et al. Influence of parental Dura and Pisifera genetic origins on oil palm fruit set ratio and yield components in their D × P progenies. *Agronomy*. 2020;10:1793.
8. Darkwah DO, Blay E, Amoatey H, et al. Genetic diversity and selection within natural dura oil palm accessions collected in Ghana for oil palm productivity improvement. *Biodiversitas*. 2020;21:3534–3538.
9. Norziha A, Zamri Z, Zulkifli Y, et al. Selection criteria of MPOB-Angola germplasm collection for yield improvement of the oil palm. *Oil Crop Sci*. 2024;9(1):20–28.
10. Monzon JP, Lim YL, Tenorio FA, et al. Agronomy explains large yield gaps in smallholder oil palm fields. *Agric Syst*. 2023;210:103689.
11. Okoye MN, Uguru MI, Bakoumé C, et al. Allelic variability of oil palm inter-population progenies based on simple sequence repeat markers. *J Plant Breed Crop Sci*. 2020;12(2):91–96.
12. Myint KA, Yaakub Z, Rafi MY, et al. Genetic diversity assessment of MPOB-Senegal oil palm germplasm using microsatellite markers. *Biomed Res Int*. 2021;2021:6620645.
13. Gan ST, Teo CJ, Manirasa S, et al. Assessment of genetic diversity and population structure of oil palm (*Elaeis guineensis* Jacq.) field genebank: A step towards molecular-assisted germplasm conservation. *PLoS One*. 2021;16(7):e0255418.
14. Zolkafli SH, Marjuni M, Abdullah N, et al. Exploring diversity in African oil palm (*Elaeis guineensis* Jacq.) germplasm populations via genotyping-by-sequencing. *Genet Resour Crop Evol*. 2025;72(5):6111–6127.
15. Sarimana U, Herrero J, Erika P, et al. Analysis of genetic diversity and discrimination of oil palm DxP populations based on the origins of Pisifera elite parents. *Breed Sci*. 2021;71(2):134–143.
16. Ithnin M, Vu WT, Shin M-G, et al. Genomic diversity and genome-wide association analysis related to yield and fatty acid composition of wild American oil palm. *Plant Sci*. 2021;304:110731.
17. Seyum EG, Bille NH, Abteu WG, et al. Genome properties of key oil palm (*Elaeis guineensis* Jacq.) breeding populations. *J Appl Genet*. 2022;63:633–650.
18. Low ETL, Chan KL, Mohd Zaki N, et al. Chromosome-scale *Elaeis guineensis* and *E. oleifera* assemblies: comparative genomics of oil palm and other Arecaceae. *G3 (Bethesda)*. 2024;14(9):jkae135.
19. Mohd Shaha FR, Liew PL, Qamaruz Zaman F, et al. Genotyping by sequencing for the construction of oil palm (*Elaeis guineensis* Jacq.) genetic linkage map and mapping of yield related quantitative trait loci. *PeerJ*. 2024;12:e16570.
20. Yaakub Z, Katialisa K, Rajinder S, et al. An integrated linkage map of interspecific backcross 2 populations reveals QTLs associated with fatty acid composition and vegetative parameters influencing compactness in oil palm. *BMC Plant Biol*. 2020;20:356.
21. Samsudin AI, Vetaryan S, Ahmad Mohktar MA, et al. Genome-wide association study for compactness traits in oil palm (*Elaeis guineensis* Jacq.). *Euphytica*. 2025;221:57.
22. Tupaz-Vera A, Ayala-Díaz I, Barrera CF, et al. Genetic gains for obtaining improved progenies of oil palm in Colombia. *Euphytica*. 2023;219:38.
23. Hermanto A, Gan SH, Mustopa IR, et al. Use of multiseasonal oil palm yield data to assess drought tolerance. *Sci Hortic*. 2023;308:111603.
24. Rachman F, Sujadi, Siregar HA, et al. The insignificance of genotype-by-environment interaction suggests a stable bunch component trait in oil palm progenies. *Menara Perkebunan*. 2026;94(1):33–46.
25. Jarquin D, Howard R, Crossa J, et al. Genomic prediction enhanced sparse testing for multi-environment trials. *G3 (Bethesda)*. 2020;10:2725–2739.
26. Pour-Aboughadareh A, Khalili M, Pocza P, et al. Stability indices to deciphering the genotype-by-environment interaction effect. *Plants*. 2022;11:414.
27. Lasky JR, Josephs EB, Morris GP. Genotype–environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell*. 2023;35(1):125–138.
28. Moeinade S, Kusmec A, Hu G, et al. Multi-trait genomic selection methods for crop improvement. *Genetics*. 2020;215(4):931–945.
29. Olivoto T, Nardino M. MGIDI: toward an effective multivariate selection in biological experiments. *Bioinformatics*. 2021;37(10):1383–1389.
30. Prathap V, Pedapati A, Ravichandran G, et al. Natural diversity in fatty acid composition and nutritional quality of African oil palm germplasm reveals opportunities for nutritional trait breeding. *Front Plant Sci*. 2026;17:1899181.
31. Bhagya HP, Kalyana Babu B, Mathur RK, et al. Oil palm (*Elaeis guineensis* Jacq.) germplasm genome-wide association analysis for the oil yield traits utilizing microsatellite markers. *Ind Crops Prod*. 2024;218:118934.
32. Shahi D, Jia G, Sumit P, et al. Multi-trait genomic prediction using in-season physiological parameters increases prediction accuracy of complex traits in US wheat. *BMC Genomics*. 2022;23:298.
33. Muvunyi BP, Zou W, Zhan J, et al. Multi-trait genomic prediction models enhance the predictive ability of grain trace elements in rice. *Front Genet*. 2022;13:883853.
34. Moreira FF, Oliveira HR, Volenec JJ, et al. Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. *Front Plant Sci*. 2020;11:681.
35. Tchounke B, Sanchez L, Bell JM, et al. Mate selection: A useful approach to maximize genetic gain and control inbreeding in genomic and conventional oil palm hybrid breeding. *PLoS Comput Biol*. 2023;19:e1010290.
36. Alemu A, Åstrand J, Montesinos-López OA, et al. Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol Plant*. 2024;17(4):552–578.
37. Krishnappa G, Savadi S, Tyagi BS, et al. Integrated genomic selection for rapid improvement of crops. *Genomics*. 2021;113(3):1070–1086.

38. Yang W, Hui F, Xuehai Z, et al. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol Plant*. 2020;13:187–214.
39. Watt M, Fiorani F, Usadel B, et al. Phenotyping: New windows into the plant for breeders. *Annu Rev Plant Biol*. 2020;71:689–712.
40. Song P, Jinglu W, Xinyu G, et al. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *Crop J*. 2021;9:633–645.
41. Ninomiya S. High-throughput field crop phenotyping: current status and challenges. *Breed Sci*. 2022;72(1):3–18.
42. Jin S, Xiliang S, Fangfang W, et al. LiDAR sheds new light on plant phenomics for plant breeding and management: Recent advances and future prospects. *ISPRS J Photogramm Remote Sens*. 2021;171:202–223.
43. Jangra S, Chaudhary V, Yadav RC, et al. High-throughput phenotyping: A platform to accelerate crop improvement. *Phenomics*. 2021;1:31–53.
44. Xiao Q, Xiulin B, Chu Z, et al. Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *J Adv Res*. 2022;35:215–230.
45. Zhang M, Shengyong X, Yutong H, et al. High-throughput horticultural phenomics: The history, recent advances and new prospects. *Comput Electron Agric*. 2023;213:108265.
46. Thompson AL, Thorp KR, Conley MM, et al. A data workflow to support plant breeding decisions from a terrestrial field-based high-throughput plant phenotyping system. *Plant Methods*. 2020;16:97.
47. Kim JY. Roadmap to high throughput phenotyping for plant breeding. *J Biosyst Eng*. 2020;45:43–55.
48. Shin MG, Ithnin M, Vu WT, et al. Association mapping analysis of oil palm interspecific hybrid populations and predicting phenotypic values via machine learning algorithms. *Plant Breed*. 2021;140:1150–1165.
49. Montesinos-López OA, Abelardo MAL, Paulino PR, et al. A review of deep learning applications for genomic selection. *BMC Genomics*. 2021;22:19.
50. Farooq M, van Dijk ADJ, Nijveen H, et al. Genomic prediction in plants: opportunities for ensemble machine learning based approaches. *F1000Res*. 2023;11:802.
51. Jeon D, Yuna K, Solji L, et al. Digitalizing breeding in plants: A new trend of next-generation breeding based on genomic prediction. *Front Plant Sci*. 2023;14:1092584.
52. Nugroho RAP, Zaag I, Lamade E, et al. Metabolomics-assisted breeding in oil palm: Potential and current perspectives. *Int J Mol Sci*. 2024;25:9833.
53. Pott DM, Durán-Soria S, Osorio S, et al. Combining metabolomic and transcriptomic approaches to assess and improve crop quality traits. *CABI Agric Biosci*. 2021;2:1.
54. Zhang R, Zhang C, Yu C, et al. Integration of multi-omics technologies for crop improvement: Status and prospects. *Front Bioinform*. 2022;2:1027457.
55. Jaramillo-Botero A, Colorado J, Quimbaya M, et al. The ÓMICAS alliance, an international research program on multi-omics for crop breeding optimization. *Front Plant Sci*. 2022;13:992663.
56. Astari RP, Basyuni M, Siregar LAM, et al. Genotypic effects on accelerated propagation of oil palm breeding materials selected using somatic embryogenesis. *Oil Crop Sci*. 2024;9(2):111–120.
57. Vu TV, Das S, Hensel G, et al. Genome editing and beyond: what does it mean for the future of plant breeding? *Planta*. 2022;255:130.
58. Sommer RJ. Phenotypic plasticity: From theory and genetics to current and future challenges. *Genetics*. 2020;215(1):1–13.
59. Budhlakoti N, Kushwaha AK, Rai A, et al. Genomic selection: A tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Front Genet*. 2022;13:832153.
60. Cock JH, Ayala-Diaz IM, Romero HM. A scheme for distribution of genetically improved oil palm plants well suited to local conditions. *Agric Syst*. 2023;212:103756.
61. Garzón-Martínez GA, Osorio-Guarín JA, Moreno LP, et al. Genomic selection for morphological and yield-related traits using genome-wide SNPs in oil palm. *Mol Breed*. 2022;42(12):71.
62. Tong H, Nikoloski Z. Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data. *J Plant Physiol*. 2021;257:153354.
63. van Dijk ADJ, Kootstra G, Kruijer W, et al. Machine learning in plant science and plant breeding. *iScience*. 2021;24:101890.