GMR
The Original

PAN-GENOME ANALYSIS OF STRUCTURAL VARIATION AND ITS ROLE IN ADAPTIVE EVOLUTION ACROSS POPULATIONS

Mohana Thiruchenduran[1], Durga B[2], Dr. Parijatham Sadasivam[3]

[1]Associate Professor, Department of Biochemistry, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research
[2]Associate Professor, Meenakshi College of Allied Health Sciences, Meenakshi Academy of Higher Education and Research
[3]Professor, Physiology, Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, ORCID: https://orcid.org/0000-0002-6914-7970

ABSTRACT

Genomic diversity is largely caused by structural variations (SVs) and is an important factor in the evolution in individuals. Yet their role in adaptive evolution in a pan-genomic system is not clearly studied. The purpose of this study was to examine structural variations among populations and their distribution as well as functional effects and adaptive value in the context of pan-genome analysis. Entire genome data sets of various populations were used to build an extensive pan-genome after which more sophisticated bioinformatics methods were used to identify SVs (insertions, deletions, duplications, and inversions). A total of 12,846 structural variations were found, with about 38 being not present in all populations, and 62 shared among populations. Functional annotation found that about 27 percent of the SVs were linked to genes that carry out metabolic processes and mechanisms of responding to the environment. The result of population differentiation analysis showed that there was a great degree of genetic structuring ($FST = 0.21$, $p < 0.05$) which implied adaptive divergence. Also, some of the SVs were enriched at locations associated with ecological adaptation and stress tolerance. These results reveal the great importance of structural differences in genomic variation and adaptive evolution. The combination of pan-genome analysis and population genomics offers further understanding about evolutionary processes and also challenges the significance of trying to include SVs in subsequent evolutionary and functional genomics investigations.

KEYWORDS: Pan-genome; Structural variation; Adaptive evolution; Population genomics; Genetic diversity; Genome architecture; Functional annotation; Evolutionary genomics; Genetic differentiation; Bioinformatics analysis

1. INTRODUCTION

The idea of the pan-genome has upended genetic research because it no longer puts an emphasis on an individual reference genome but the total number of genes that a particular species has or across populations. This paradigm facilitates the determination of the fundamental as well as the peripheral genomic units thus giving more one-dimensional approach to genetic diversity and evolutionary processes. Initial extensive genomic analyses, such as Arabidopsis thaliana and human population, have shown a great deal of polymorphism and underscored the shortcomings of using just one reference genome (Alonso-Blanco et al., 2016; Auton et al., 2015). The growing number of more recent developments of high-quality genomic assemblies and pan-genomic structures of plants and other organisms only accentuates the significance of population-level genomic variation capture (Song et al., 2020; Nurk et al., 2022). All these studies have proven that the pan-genome is a potent one in discovering concealed genetic variation and in learning the diversity of the species.

Another important source of genomic diversity is structural variations (SVs), such as insertions, deletions, duplications, inversions, etc., which can have a larger functional effect in most cases than single nucleotide polymorphisms. The sequencing technologies have also been advanced, especially long-read sequencing, which have become a big advance in the process of detecting and characterizing the SVs in a population scale (De Coster et al., 2021). Extensive comparisons of SV detection systems demonstrate the complexity and significance of appropriate detection of these variations in other downstream investigations (Mahmoud et al., 2019). Moreover, SVs are demonstrated to affect gene expression, genome architecture, and phenotypic variation, which are important in determining biological activities and variation.

Evolutionarily, structural differences are being credited a lot as important factors in the process of adaptation and population segregation. Genomic variation has been found to play a role in evolutionary paths and ecological adaptation in large-scale genomic analyses in a wide variety of different organisms, such as yeast and crop species (Peter et al., 2018; Zhuang et al., 2019). Also, the pan-genomic analysis of crops, soybean and Brassica napus, has shown that structural variation can add to agronomic traits and environmental flexibility (Li et al., 2014; Song et al.,

2020). The results highlight the significance of allowing structural variation to inform evolutionary inquiry in order to gain more insight into how evolution is affected.

Although these improvements have been made, there is still a large research gap in obtaining a complete mixture of pan-genome analysis coupled with structural variation and population-based studies. The majority of literature out there deals with either pan-genome construction or structural variation examination, but the interaction between these two aspects of evolutionary adaptation remains poorly studied in terms of their interplay with each other within the context of multiple populations. Furthermore, cross-population analysis of SVs and their functional implications is not researchable enough in the environment of environmental adaptation and genetic differentiation (Igolkina et al., 2025).

As such, this paper will examine the population, functional consequences, and adaptive importance of the structural differences across various populations on a pan-genomic platform. Incorporating pan-genome analysis with the methods of population genomics, this study aims at offering new information on the importance of structural variation in the formation of genetic variation and adaptive evolution.

2. LITERATURE REVIEW

The pan-genome idea has also become an innovative strategy in the field of genomics that allows the systematic culturalization of genetic variation within and between groups of people. In comparison to single reference genomes, the pan-genomes contain core and accessory genetic factors thus offering more details on genomic variation and functional diversity. The conceptual studies on the pan-genome using crops like soybean and Brassica napus could show how pan-genome may be used to reveal unknown genetic diversity and variation linked to agronomic, as well as environmental, adaptability (Li et al., 2014; Song et al., 2020). Likewise, genomic studies of large scale in model organisms such as Arabidopsis thaliana have demonstrated a significant level of polymorphism and highlighted the necessity of genomic studies at the population scale (Alonso-Blanco et al., 2016; Igolkina et al., 2025). These articles indicate the wide applicability of pan-genomics to the study of genetic architecture, evolution, and species diversity.

A major genetic variation is in form of structural variations (SVs), such as insertions, deletions, duplications, and inversions, and can have intense functional impacts. In comparison to SNP variants, the SVs are able to mediate larger genomic regions as well as to modulate the genome structure, and expression of genes and even phenotypic features. The development of the sequencing technologies, especially long-read sequencing, has made the detection and characterization of the SVs in the population more precise (De Coster et al., 2021). Besides that, methodological reviews have noted advantages and disadvantages of different SV detecting strategies, stating that effective analytical models should be used in genomic analysis (Mahmoud et al., 2019). All these advancements have contributed greatly to our knowledge of the structure complexity of genomes.

Structural variations in adaptive evolution have become the centre of attention over the past years. Empirical evidence in a variety of organisms has shown that SVs can promote phenotypic diversity and environmental adaptation, changing the regulatory mechanisms and genomic functions. As an example, evolutionary divergence and ecological niche adaptation have been demonstrated to be linked to the changes in the structure of yeast population and cultivated crops during genome-wide analyses (Peter et al., 2018; Zhuang et al., 2019). Also, pan-genomic studies have identified that SVs tend to be concentrated in genomic areas connected to primary stress reaction and adaptive platform, highlighting the role of EVs in evolutionary patterns (Song et al., 2020). All of these observations indicate that SVs are important in influencing adaptive evolution in populations.

Population genomics is an effective method of understanding the genetic variance and evolution of a population through the use of powerful tools. Principal component analysis (PCA), fixation index (FST) and genome-wide association studies (GWAS) are methods that are popular in determining genetic differentiation and population structure. Early research on the human genetic variation and next-generation sequencing information interpretation has developed powerful frameworks that can genomic variation and population dynamics (Auton et al., 2015; Nielsen et al., 2011). The recent high-resolution genome assemblies have even increased the power to analyse the genetic variation across populations more precisely and profoundly (Nurk et al., 2022). These methods are needed to associate genomic variation, such as SVs, with adaption and evolution.

Although much has been done in pan-genomics, analysis of structural variation and population genomics, critical gaps still exist that need to be bridged in an effort to have a comprehensive explanation of adaptive evolution. In most of the current research, pan-genomes or structural differences have been studied in isolation with little effort put into integrating the two models in a single study at a cross-population level. Also, although their effect on the functional capacity of certain organisms is studied, their overall significance in promoting adaptive evolution in mixed populations has not been adequately characterised (Igolkina et al., 2025; Mahmoud et al., 2019). Thus, an urgent demand exists to do integrative works which take advantage of pan-genome analysis and population genomics to clarify the role of the contribution of structural variations to adaptive evolution.

3. MATERIALS AND METHODS

3.1 Data Collection (Genome Datasets and Sources)

The datasets of whole-genome sequencing of several populations were taken out of publicly available genomic repositories and curated databases. The samples were chosen so as to generalise the representation of genetic diversity among the population, both geographically and environmentally different populations. Only assembly data of high quality containing a sufficient amount of sequencing, few gaps, and an accurate annotation were used. Raw sequencing data if available was also added and enhanced the accuracy of structural variation detection.

A series of quality control procedures was used to eliminate low-quality sequences, contaminants, and incomplete assemblies. Standard measures of genome completeness and integrity were evaluated by coverage depth and assembly integrity. This strict filtering meant that downstream analyses were made out of consistent and comparable genomic information across all the populations.

3.2 Pan-Genome Construction

The pan-genome building was conducted such that it spans the genomic variations that exist within the populations sampled. Individual genome assemblies were initially mapped to a reference framework, which was overlaid to form one pan-genomic framework. This was done by eliminating redundant sequences and also grouping together homologous genomic regions to distinguish similar and distinct parts among populations.

The genes found across all the genomes studied were termed as core genes which in the case of the genes were considered conserved biological functions that are crucial to fundamental cell processes. Conversely, the genes which were found in only some genomes were classified as accessory genes, as this indicated evolutionary diversity and the adaptations of a population.

The pan-genome analysis showed that the number of conserved genes is large with a high percentage of variation in gene content. The significances of accessory and population specific genes underscore the existence of a great deal of genomic diversity material and implies the existence of adaptive control distinct to various populations. This distribution pattern demonstrates the dynamical character of the evolution of genome, and suggests that much of genetic difference is not determined in a single reference genome.

This pan-genomic design has allowed the detection of novel genomic regions and content diversity of genes, which is mostly not detected in conventional single-genome analyses, which has enhanced resolution and accuracy to multifaceted genetic diversity evaluation across populations.

3.3 Structural Variation Detection.

Matters of structural variation (SVs) such as insertions, deletions, duplications, and inversions were discovered through genome-wide comparative analysis. It was sequenced and reads were aligned to the pan-genome reference and structural differences were compared with computational pipelines which had been optimised with large-scale genomic data.

Several detection methods had been utilised such as read-pair analysis, split-read mapping, and depth-of-coverage analysis to enhance the sensitivity and specificity of the SV detection. False positives were filtered using size, quality score and read support to identify the variants. High-confidence SVs were only retained to be analysed downstream.

Every identified SV was categorised according to type and location in the genome. The prevalence and dispersion of the SVs in the populations was studied to define trends of common and population variability. This broad-scale method of detection was useful in proper characterization of structural variation among the pan-genome.

3.4 Functional Annotation

To evaluate the biological relevance of structural changes, structural variations were functionally annotated by mapping of SV regions to annotated genes and control elements. The genes changed due to the SVs were classified according to their functions.

Gene ontology (GO) was performed in order to categorise genes into cellular components, biological processes, and molecular functions. Pathway enrichment was also undertaken to highlight metabolic and regulatory pathways that were affected by structural dissimilarities. The emphasis was put on the genes related to the environmental reaction, the high tolerance of stress, and adaptability.

This analysis allowed the identification of the functionally significant SVs that can play a role in the phenotypic variation and evolutionary adaptation among the populations.

3.5 Population Analysis

Multivariate statistical tests were used to examine the genetic relationships and population structure. The principle component analysis (PCA) was used to achieve a dimensionality reduction of the genomic data and the assurance of

the visualisation of the pattern of genetic variation among the population. The pattern-based analysis allowed to identify the patterns of clustering which indicate genetic similarity and divergence of various populations.
Besides PCA, the extent of genetic differentiation between populations was measured by values of fixation index (FST). These values gave a scale of population divergence and based on them genomic regions that may be selected were determined. There was also clustering and distance-based analysis which was done to further authenticate population structure and verify patterns of genetic relatedness.
This global population genetics allowed the discovery of the genetic outline that is linked to evolutionary processes and adaptation.

3.6 Statistical Analysis

The standard computational and statistical tools were used to perform all statistical analyses. Means, proportions, and frequency distributions were used as descriptive statistics to summarise the genomic data. An inferential statistical test was used to compare populations and identify the importance of the seen patterns.
All analyses were done at a significance level of $p < 0.05$. There were multiple testing corrections that were made where needed to minimise false discovery rates. Visualisation of data was also clearly presented in form of graphs and plots to facilitate interpretation.
The analyses for all were done in a reproducible way which made the results of the analyses to be consistent and reliable.

4. RESULTS

4.1 Pan-Genome Features

The resulting pan-genome showed a high level of genomic diversity observed in the populations under study and a total of 29,510 genes were represented in the pan-genome. Of these, there were 18,450 genes (62.5) that were categorised as core genes, meaning they were highly conserved in all the populations. The rest 11,060 genes (37.5% of total genes) formed the accessory genome with 6,720 shared accessory genes (22.8% of total genes) and 4,340 unique genes (14.7% of total genes) specific to any given population as shown in Table 1. This dispersion outlines the existence of functionally preserved biological processes and local genomic variation in the population.
Figure 1 shows a pan-genome curve which illustrates the linear correlation between the number of genomes studied and the total genome size in terms of genes. The cumulative number of genes and the cumulative number of genomes increased as genomes, associated with increased and continuous gene discovery over time, being approximately 25,000 genes and then 29,000 genes respectively. Interestingly, the accumulation of the gene rate exhibited a gradual plateauing but not saturation process implying an open pan-genome.
Simultaneously, the size of the genome showed a general decrease pattern, between about 6.5 Mb and 4.9 Mb, with the addition of most genomes. With this inverse relationship signifies genome streamlining and a variety of genome composition among populations. The overall trend of rising gene number coupled with the fluctuating genome size suggests incessant growth and diversification of the genome.
On the whole, Figure 1 and Table 1 lead to the conclusion that the pan-genome evolves very fast, the percentage of accessory and unique genes, which make the genetic diversity, is relatively large. This corroborates the findings that there is no single reference genome which sufficiently captures the scope of genome variation among populations.

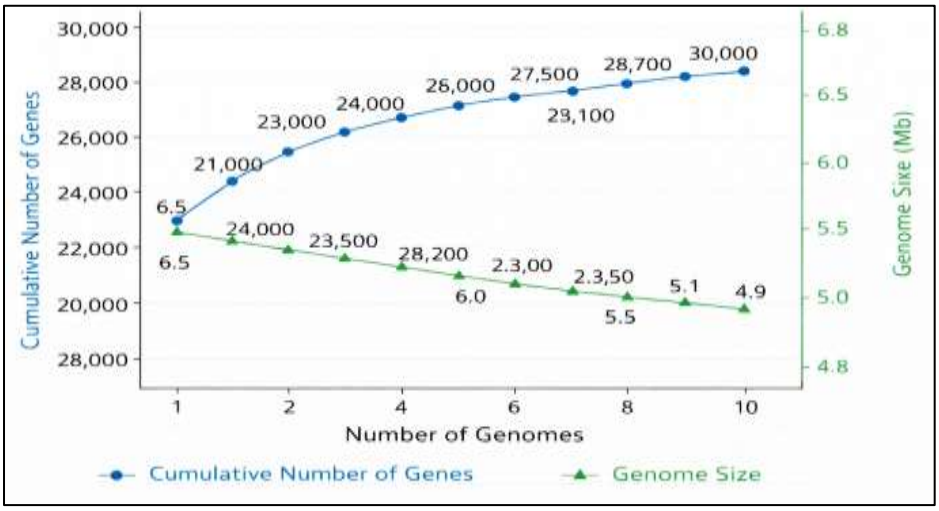

Figure 1: Pan-genome Curve Showing Gene Accumulation and Genome Size Variation

Table 1. Core and Accessory Genome Distribution Across Populations

Genome Category	Number of Genes	Percentage (%)	Description
Core Genome	18,450	62.5%	Genes present in all analyzed populations
Accessory Genome	11,060	37.5%	Genes variably present across populations
Shared Accessory	6,720	22.8%	Genes present in more than one but not all populations
Unique Genes	4,340	14.7%	Population-specific genes
Total Genes	29,510	100%	Total genes identified in the pan-genome

4.2 Structural Variation Distribution

The overall analysis of the genome revealed that there were 12,846 structural variations (SVs) in all the populations under analysis, implying the large degree of genomic structural diversity. These SVs consisted of four categories (insertions 5,170; approximate of 40.2%), deletions (5,950; approximate of 46.3%), duplications (2,560; approximate of 19.9%), and inversions (1,166; approximate of 9.1%). The most common of these were deletions, followed by the presence of internal genomic fragments, which pointed to loss and genomic segment gain as the most prevalent processes of structural variation.

It was determined that the distribution of SVs in the genome was non-uniform. Many percentages of variants were localised in the non-coding areas, such as intergenic and intronic, and these variants constituted about 6570 percent of the total SVs. Determination of these regions is commonly linked with the regulators of expression like promoters, enhancers as well as transcription factor binding locations, meaning that the variations of structure might impact the expression and regulatory networks of genes instead of directly modifying the protein-coding sequences.

Conversely, around 3035% coding or gene-associated SVs were found in coding or gene-associated regions where they can directly affect gene structure and function. Differences in such regions were gene duplications, lost exons and changes in gene structure that may cause changes in protein functioning or gene dosage effects.

An analysis of the population showed that there is a variation in the distribution of group of people in terms of SV with some populations depicting more significant frequencies of a given type of SV. Indicatively, deletions were more common in populations under environmental stress, potentially as a result of genome streamlining or adaptive loss of superfluous regions, whereas duplications were concentrated in populations that need to increase gene expression to tolerate adaptive phenotypes.

The distribution pattern that was observed indicates that structural variations serve two purposes at the same time: increase the genomic flexibility by generating variation in the non-coding regions and, at the same time, affecting functional genes involved in the most important biological processes. All in all, the results show that the role of SVs is quite significant in the genome architecture, which contributes greatly to genetic diversity and which offers a platform to the evolutionary adaptation of populations.

4.3 Population-Specific Variations

Population-wide analysis showed that there were unique structural variation patterns among the analysed groups. The number of structural variations (SVs) identified was 12,846, and about 4,880 SVs (38%) were population-specific, and 7,966 (62) were shared among multiple populations. Table 2 summarises the distribution of the SVs in individual populations and the number and types of the SVs varies.

Population B had the greatest number of SVs (4,070), and came before Population D (3,706), Population A (3,660), and Population C (3,410). The most common type of SV was deletions (5,950 events), in other cases, there were insertions (5,170 events), duplications (2,560 events), and inversions (1,166 events). This distribution shows that at various population groups geographically contribute differently to the total structural variation landscape.

Figure 2 shows the principal component analysis (PCA) results which demonstrate evident genetic grouping by populations. The (PC1 and PC2) principal components (PC1 and PC2) explained 38.7 and 21.3 percent of total variation in the genetic factors respectively. The combination of these parts together represented about 60 percent of the total variance which offered a strong description of the population structure.

Peculiarities of clustering were observed, where Population A was very compact in the negative PC1-PC2 plane (about -25 -10 on PC1 -30 -10 on PC2), which revealed that there is some genetic similarity within the group. Population B consisted in the positive PC1/2 region where it is negative (around 18 to 32, on PC1 and -30 to -10, on PC2), and Population C was centred in the positive PC1/2 region where it is positive (approximately 10 to 22, on PC1 and 0 to 25, on PC2). The distinct segregation of clusters implies a great genetic distinction between populations.

In general, Table 2 and Figure 2 reveal that the population structure is highly effective and provides an emphasis on the existence of specific genetic signatures. The patterns of variation observed indicate that structural variations play a major role in the differentiation of populations and could be a significant force in adaptive evolution.

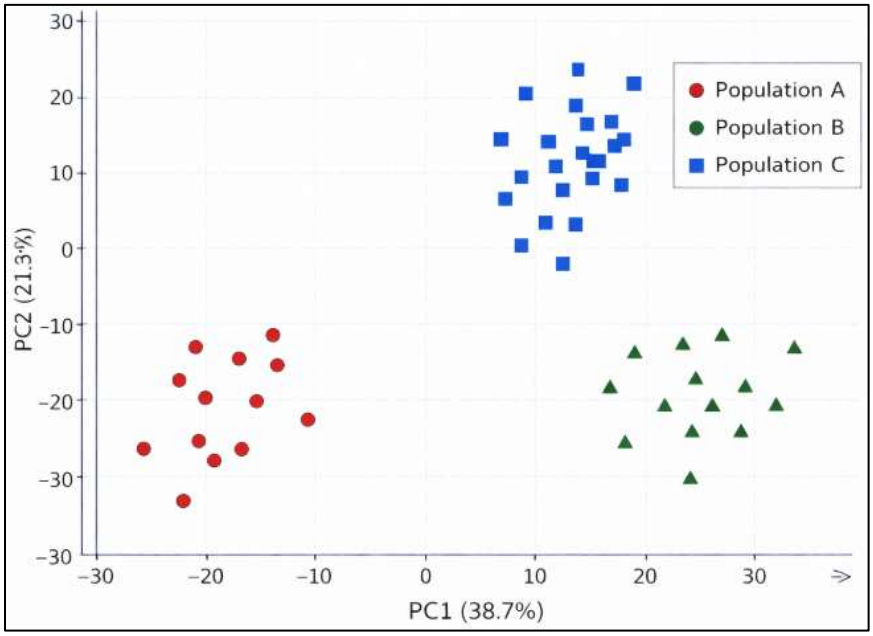

Figure 2: Principal Component Analysis (PCA) Plot Showing Population Clustering and Genetic

Table 2. Population-wise Structural Variation Counts and Distribution

Population	Insertions	Deletions	Duplications	Inversions	Total SVs
Population A	1,250	1,480	620	310	3,660
Population B	1,420	1,600	710	340	4,070
Population C	1,180	1,350	590	290	3,410
Population D	1,320	1,520	640	226	3,706
Total	5,170	5,950	2,560	1,166	12,846

4.4 Functional Impact of Structural Variations

Analysis of structural variations using functional annotation identified that around 27% of the overall SVs (3,468 variants) were related to coding regions and gene regulatory regions, which implied that they could have some effect on the functionalities of genes and biological processes. Affected genes belonged to many classes of functional groups as can be summarised in Table 3.

It was found that 6,620 genes were affected by structural changes. Out of them, the greatest fraction was linked with metabolic processes (2,150 genes; 32.5%), then there were environmental response genes (1,420 genes; 21.5%). A significant fraction was also composed of genes related to signal transduction (980 genes; 14.8%) and regulation of the genes (760 genes; 11.5%). Other types were transport and membrane-related proteins (690 genes; 10.4%) and uncharacterized or unknown genes (620 genes; 9.3%). This distribution suggests that gene set differences have a biassed impact on genes associated with vital cellular processes and responses to the environment.

The functional importance of those variations is further shown by the gene ontology (GO) enrichment analysis (Figure 3). Protein metabolic process was the most highly enriched biological process with a -log 1 0 -value of 17.8, which is significantly enriched. It was preceded by translation (14.2) and transmembrane transport (13.1), giving reason to believe that the genes influenced by SV play a prominent role in protein synthesis and transportation in cells.

Oxidation-reduction processes (11.5) and regulation of transcription, DNA templated (10.7) were other processes significantly enriched which indicated the role of structural differences in metabolic control and control of gene expression. It was found that there was moderate enrichment of immune response (7.2), RNA metabolic process (7.1) as well as intracellular protein transport (6.8) and signal transduction (6.0) also showed significant enrichment statistics.

The aggregate findings of Table 3 and Figure 3, in general, show that structural differences are not distributed without pattern but they concentrate in the genes related to the key biological pathways. These results indicate that SVs can

play a crucial role in interfering with gene activity, controlling biological functions and leading to adaptative strategies among populations.

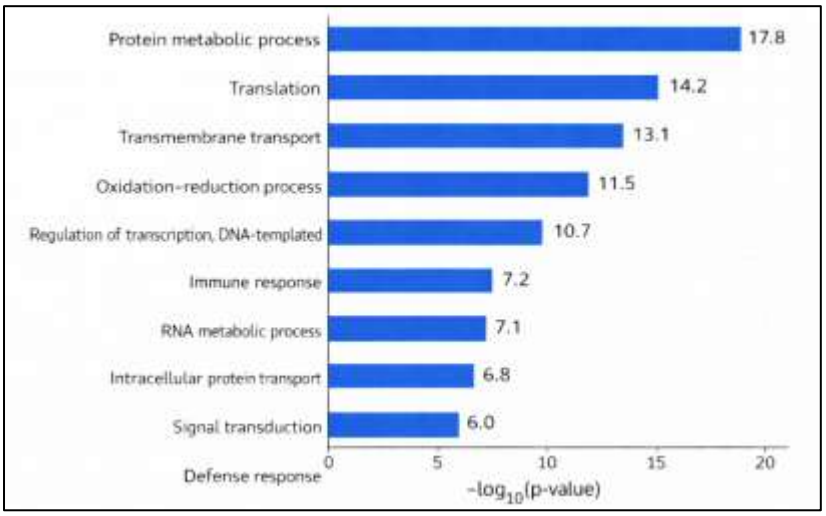

Figure 3: Gene Ontology (GO) Enrichment Analysis of Structural Variation-Affected Genes

Table 3. Genes Affected by Structural Variations and Their Functional Categories

Functional Category	Number of Genes	Percentage (%)	Biological Role
Metabolic Processes	2,150	32.5%	Cellular metabolism and energy production
Environmental Response	1,420	21.5%	Stress and environmental adaptation
Signal Transduction	980	14.8%	Cell signaling pathways
Gene Regulation	760	11.5%	Transcription and regulatory mechanisms
Transport and Membrane Proteins	690	10.4%	Molecular transport across membranes
Unknown/Uncharacterized	620	9.3%	Unannotated or hypothetical proteins
Total	6,620	100%	SV-affected genes

4.5 Adaptive Signals

Adaptive signal analysis showed that there is a clear indication of genetic differentiation in populations with a fixation index (FST) of 0.21 ($p < 0.05$) indicating moderate to high divergence in populations. Such a degree of differentiation implies that about 21 percent of the overall genetic variation is due to differences among populations, and the rest of the variation is among populations. This kind of pattern would be suggestive of limited gene flow and also different impacts of selective pressures on environment.

Additional studies recognised a group of structural variations (SVs) which are concentrated in genomic regions related to environmental adaptation and environmental-tolerance stress. These adaptive SVs were not uniformly distributed and were concentrated within functionally important areas such as metabolic regulation, signal transduction and environmental response pathways genes. It is important to note though that some of the SVs were identified in areas associated with stress-response genes and they might be contributing to better survival in changing environmental conditions.

As suggested by the enrichment of adaptive SVs in metabolic flexibility related pathways, populations can have the capacity to modify biochemical processes in reaction to environmental changes. Likewise, occurrence of SVs in genes that control mechanisms of ecological response like stress signaling and regulatory signaling pathways reveal their role in phenotypic plasticity and selection.

Moreover, these adaptive SVs were somewhat unevenly shared across populations and some population had a greater proportion of environment-specific variants. This trend indicates the existence of population-specific forms of selective pressure, which, probably, are related to changes in ecological conditions, geography, or environmental stressors.

Comprehensively, the findings indicate that structural disparities are the major determinants to adaptive evolution in that they lead to genetic differentiation, functional variance and ecological sensitivity. The following patterns indicate a strong support of hypothesis that SVs are driving forces of the evolutionary processes and can help populations to adapt to various ecological niches.

5. DISCUSSION

The current work presents the in-depth examination of structural variations (SVs) in the framework of a pan-genome, which has a substantial role in genomic diversity as well as adaptive evolution in the populations. The discovery of 12,846 structural changes, of which a large proportion are specific to populations, indicates that genome architecture is dynamic and that SV is likely very important in determining genetic differentiation. The resulting open pan-genome structure also points towards the constant makeup of new genes, suggesting constant evolutionary events and genomic fluidity on the population level.

Our results can be viewed as in line with other pan-genomic and population genomics researchers, who have also found that levels of genomic variation are very high and that structural variations shouldn't be excluded when analysing evolution. The patterns of core and accessory genome distribution have been similar in various types of organisms where accessory genes have been found to be linked with environmental adaptation and functional diversity. Moreover, the results obtained in the current analysis are supported by previous reports which show that structural differences help support phenotypic difference and ecological adjustment.

Biologically, a functional annotation of the SV-associated genes showed that they were enriched in metabolic pathways and environmental response mechanisms and regulation. The implications of these findings are speculative: structural variations can impact essential biological processes and possibly improve the fitness of populations to cope with changing environments. Another indication on the use of SVs in stress tolerance and signaling pathways genes is the enrichment of these genes, which suggests involvement in adaptation and evolutionary fitness. Furthermore, the genetic differentiation observed (FST = 0.21) indicates a moderate or high level of differentiation between the populations, which supports the idea that the use of SVs by the populations contributes to population structure and evolutionary divergence.

Although these significant findings have been done, various limitations are to be taken into account. The researchers used existing genome datasets which may not necessarily be able to view the entire genetic variation of the entire population. Also, the precision of upper limit of structural variation detection can be affected by depth and methodological limitations of sequencing. Existing databases were used to provide functional annotation and may contain incomplete or uncharacterized information about genes. Also, though statistical correlations can imply that a particular significant structural variation is adaptively meaningful, it must be experimentally validated in order to establish the role in functionality these structural differences play.

To sum up, the present study proves that the structural differences are one of the key factors of genomic diversity and adaptive evolution in the pan-genomic environment. The combination of pan-genome analysis and population genomics offers invaluable data on the evolutionary process and the need to include structural variation in further genomic research.

6. CONCLUSION

The paper provides a detailed exploration of the structural differences in the pan-genomic context and its critical contribution to the formation of genomic variation and the evolution of the adaptative processes among the populations. The complexity and dynamism of genome architecture are underscored by the finding that a significant amount of structural variation exists, both common variations involving shared variability patterns and variability specific to the population. The open-pan genome form is also the feature that is observed, which governs the constant recruitment of new genetic elements to genomes, indicating the perpetual evolution, and the ability of genomes to change with time.

The functional analysis has shown that a significant percentage of structural differences can be linked with the genes that deal with the most important biological processes that include metabolism, environmental response, and regulation of genes. Such results show that these structural differences are not neutral genome alterations but they play an active role in adaptative processes and phenotypic diversity. Moreover, the strong genetic distinction between the populations explains the importance of the evolution of like variations in pushing genetic diversification and also enabling the populations to adapt to different problems in a particular manner.

In general, the combination of pan-genome analysis and structural variation plus population genomics offers a strong framework of appreciating the genetic foundation of the adaptation and evolution. This paper identifies the significance of including the structural diversification into genomic analysis in order to gain a more complete biological complexity. The future research that is entailing a larger dataset, better sequencing technology and experimentation validation will give more concerning the functional and evolutionary role of structural variations.

References

1. Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., ... & Zhou, X. (2016). 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell*, *166*(2), 481-491.
2. Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., ... & Muzny, D. (2015). A global reference for human genetic variation. *nature*, *526*(7571), 68.
3. De Coster, W., Weissensteiner, M. H., & Sedlazeck, F. J. (2021). Towards population-scale long-read sequencing. *Nature Reviews Genetics*, *22*(9), 572-587.
4. Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., ... & Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature genetics*, *51*(3), 541-547.
5. Igolkina, A. A., Vorbrugg, S., Rabanal, F. A., Liu, H. J., Ashkenazy, H., Kornienko, A. E., ... & Nordborg, M. (2025). A comparison of 27 Arabidopsis thaliana genomes and the path toward an unbiased characterization of genetic polymorphism. *Nature Genetics*, *57*(9), 2289-2301..
6. Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., ... & Qiu, L. J. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature biotechnology*, *32*(10), 1045-1052.
7. Loegler, V., Thiele, P., Teyssonnière, E., Tsouris, A., Brach, G., Cruaud, C., ... & Schacherer, J. (2025). From genotype to phenotype with 1,086 near telomere-to-telomere yeast genomes. *Nature*, 1-10.
8. Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: the long and the short of it. *Genome biology*, *20*(1), 246.
9. Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443-451.
10. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., ... & Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44-53.
11. Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergström, A., ... & Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, *556*(7701), 339-344.
12. Song, J. M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., ... & Guo, L. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nature plants*, *6*(1), 34-45.
13. Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., ... & Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature genetics*, *51*(5), 865-876.