Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative genome-wide characterization of simple sequence repeats in coriander (Coriandrum sativum) and related Apiaceae

Cristian Mendoza
Facultad de Ingeniería y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas

Karina Barboza
Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria La Consulta

Andrés Morales
Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria La Consulta

Gokhan Sadik
Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University

Metin Koçak
Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University

Mehmet Alp Furan
Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University

Dariusz Grzebelus
Department of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow

Carlos I. Arbizu
Facultad de Ingeniería y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas

Mehtap Yildiz
Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University

Pablo F. Cavagnaro
cavagnaro.pablo@inta.gob.ar

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Research Article

Keywords: Coriander, Apiaceae, Simple sequence repeats, Genetic diversity, Comparative genomics

Posted Date: August 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10588820/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparative genome-wide characterization of simple sequence repeats in coriander (*Coriandrum sativum*) and related Apiaceae

Cristian Mendoza[1], Karina Barbosa[2,3], Andrés Morales[2,4], Gokhan Sadik[5], Metin Koçak[5], Mehmet Alp Furan[5], Dariusz Grzebelus[6], Carlos I. Arbizu[1,7*], Mehtap Yildiz[5]*, Pablo F. Cavagnaro[3,6,8]*

[1] Facultad de Ingeniería y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Amazonas, 01001, Perú.
[2] Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria La Consulta, La Consulta M5567, Mendoza, Argentina.
[3] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires C1425FQB, Argentina.
[4] Universidad Nacional de Cuyo, Facultad de Ciencias Agrarias; and Instituto de Biología Agrícola de Mendoza (IBAM CONICET) Laboratorio de Biología Molecular, Luján de Cuyo M5505, Mendoza, Argentina.
[5] Department of Agricultural Biotechnology, Faculty of Agriculture, Van Yuzuncu Yil University, Van 65080, Turkey.
[6] Department of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Krakow 31-425, Poland.
[7] Centro de Investigación en Germoplasma Vegetal y Mejoramiento Genético de Plantas (CIGEMP), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Chachapoyas, Perú.
[8] Instituto Nacional de Tecnología Agropecuaria (INTA) Estación Experimental Agropecuaria Mendoza, Lujan de Cuyo M5507, Mendoza, Argentina.

* Correspondence: carlos.arbizu@untrm.edu.pe (C. I. A.); mehtapyildiz@gmail.com (M. Y.); cavagnaro.pablo@inta.gob.ar (P. F. C.)

ORCID IDs

Cristian Mendoza 0000-0003-1686-7023
Karina Barboza: 0000-0003-1941-4492
Andrés Morales 0000-0003-4962-727X
Metin Koçak 0000-0002-8109-5245
Dariusz Grzebelus 0000-0001-6999-913X
Mehtap Yıldız 0000-0001-6534-5286
Carlos I. Arbizu 0000-0002-0769-5672
Pablo F. Cavagnaro 0000-0001-5838-0876

Abstract

Background Simple sequence repeats (SSRs) are among the most informative molecular markers for plant genetics, yet their genome-wide distribution and marker potential in coriander (*Coriandrum sativum*) and related Apiaceae remain poorly characterized. Here we surveyed and compared SSRs in the publicly available whole-genome sequences of five Apiaceae species: coriander, carrot (*Daucus carota*), celery (*Apium graveolens*), gotu kola (*Centella asiatica*) and fennel (*Foeniculum vulgare*).

Results A total of 569,531 perfect microsatellites were recovered in the 2,118.68 Mbp coriander assembly, an overall density of 268.8 SSRs/Mbp that places coriander at the low-density end of the family. Tetranucleotides were the most abundant repeat type in coriander, celery and fennel, while dinucleotides predominated in carrot and gotu kola; across all genomes exons were consistently depleted in SSRs and enriched in frame-preserving trinucleotides and AG-type motifs, a shift consistent with selection against frame-disrupting repeats in coding sequence. From the coriander loci we designed 1,806 primer pairs and assessed their transferability *in silico* across the family; requiring a predicted amplicon within 50 bp of the expected size, verified transferability was low and strongly dependent on the mismatch tolerance applied, ranging from 2.2 to 7.6% in carrot, 4.3 to 15.6% in celery, 2.8 to 11.0% in fennel and 0.2 to 4.2% in gotu kola across thresholds of one to three mismatches per primer, whereas a homology-only screen of the same primers returned 74.9 to 98.0%, indicating that homology screening substantially overstates cross-genus transferability. Comparison of repeat number at 552 orthologous motifs, of which 191 had sufficient locus support for analysis, showed that length change was nearly universal, with expansions and contractions roughly balanced across repeat types and stable loci essentially absent, while repeat number was positively correlated with allele-length divergence (Spearman's $\rho = 0.53$ to 0.86), pointing to replication slippage as the dominant process. Applied to 14 coriander accessions of diverse origin, six polymorphic loci resolved 26 alleles (mean 4.3 per locus; mean PIC = 0.68) and revealed limited diversity with weak population structure, in line with the narrow genetic base of the crop.

Conclusions This work delivers the first genome-wide SSR catalogue for coriander and a set of cross-genus primer pairs that can serve as practical anchors for comparative genomics, mapping and marker-assisted breeding across Apiaceae.

Keywords: Coriander, Apiaceae, Simple sequence repeats, Genetic diversity, Comparative genomics

Introduction

Coriander (*Coriandrum sativum* L.) is an annual herbaceous member of the Apiaceae family, native to the Mediterranean region and cultivated worldwide as a culinary and medicinal crop [1, 2]. Both its foliage and seeds are economically important, and the species is valued in the food, pharmaceutical and fragrance industries for its essential oils, flavonoids, phenolics and terpenoids [3]. Cultivated for several millennia and believed to be of Mediterranean origin [4], coriander is the only widely cultivated member of the genus *Coriandrum*, together with the wild species *C. tordylium* (Fenzl) Bornm. [5]. It is a diploid species (2n = 2x = 22) whose immature plants are commonly known as cilantro, whereas its seeds are referred to as coriander [6].

The Apiaceae family comprises approximately 466 genera and 3,820 species [7] and includes several crops of major agronomic importance, such as carrot, celery, fennel, parsley and dill [8-10]. Despite the economic significance of these species, genomic resources remain limited for coriander and most of its relatives within the family, restricting the development of molecular tools and slowing the pace of genetic improvement.

Among the available molecular markers, simple sequence repeats (SSRs), also known as microsatellites, are short tandemly repeated DNA motifs consisting of one to six nucleotides. They are abundant in both prokaryotic and eukaryotic genomes and contribute to genetic variation, genome evolution, and gene regulation [11-13]. Owing to their high polymorphism, co-dominant inheritance and reproducibility, SSRs have been extensively used in genetic diversity studies, population genetics, linkage mapping and marker-assisted breeding [14, 15]. Consequently, they remain widely employed across plant species for crop improvement and the assessment of genetic diversity [16, 17].

In coriander, however, only a limited number of SSR markers have been developed and applied [18]. Furthermore, the transferability of these markers to related species is generally low, constraining comparative genetic analyses and breeding applications. Although advances in next-generation sequencing have enabled rapid and cost-effective genome-wide SSR discovery in numerous crops [19, 20], information regarding the abundance and distribution of SSRs in coriander and other Apiaceae genomes remains scarce. Existing studies within the family have focused primarily on transcriptomic or organellar resources. For example, Song et al. [21] developed the Coriander Genomics Database using cilantro and carrot genomic resources, Wang et al. [20] reported the coriander mitochondrial genome, and Tulsani et al. [22] characterized the transcriptomes of 14 coriander cultivars for marker development. To our knowledge, no comprehensive genome-wide analysis of SSRs in coriander or closely related Apiaceae species has yet been reported.

A genome-wide survey is therefore needed to determine the abundance, distribution and composition of SSRs across these genomes. Moreover, a comparative analysis among related species could provide insights into microsatellite organization within the family and facilitate the identification of conserved loci with potential utility for marker transferability and comparative genomics. In this study, we characterized and compared SSRs in the publicly available whole-genome sequences of coriander, carrot, celery, gotu kola and fennel. Specifically, we aimed to (1) identify and characterize SSR loci in the coriander genome, (2) compare SSR frequency, distribution and motif composition among the five species, and (3) evaluate the utility of coriander-derived SSRs for cross-species transferability and genetic diversity analysis.

Materials and Methods

Source of whole genome sequences

The whole genome sequences of coriander, carrot (*Daucus carota* L.), celery (*Apium graveolens* L.), gotu kola (*Centella asiatica* L.) and fennel (*Foeniculum vulgare* Mill.) were used in this study to characterize the distribution of simple sequence repeats across the Apiaceae family. Detailed information on the genomic sequences used, including assembly versions, sequence sources, project IDs and total sequence lengths, is provided in Supplementary Table S1. Whole-genome SSR densities were compared across all five species, whereas the analysis of specific genomic fractions, including non-coding DNA, genes (introns + exons), untranslated genic regions (introns + UTRs) and exons, was restricted to coriander, carrot and celery, the three species whose genomes are annotated. For gotu kola and fennel, whose genome sequences are not annotated, only whole genome sequence data were analyzed.

Detection and analysis of SSRs in different fractions of genomic sequences

Although SSRs are commonly defined as repeats of 1- to 6-bp motifs [23], in this work we also analyzed repeats with 7-bp motifs, whereas mononucleotides were not considered due to the difficulty of distinguishing bona fide microsatellites from sequencing or assembly errors (these biases are more common in whole genome shotgun sequence data). Thus, DNA sequences were searched for perfect microsatellites with a basic motif of 2–7 bp using the computer program MISA (MIcroSAtellite identification tool) [24]. Repeats with a minimum length of 12 (for di- to tetranucleotides), 15 (for pentanucleotides), 18 (for hexanucleotides), and 21 (for heptanucleotides) were recorded.

SSR primer design

Oligonucleotide primers flanking SSRs identified in the coriander genome were designed using Primer3 (v. 0.4.0) software [25]. Primers were designed to generate amplicons of 100-400 bp in length (optimum 250 bp) with the following minimum, optimum, and maximum values for Primer3 parameters: primer length (bp): 18-22-27; Tm (°C): 57-60-63; GC content (%): 30-50-70. Other parameters used the program default values.

***In silico* PCR and cross-species transferability**

Cross-species transferability of the coriander primer pairs was evaluated by *in silico* PCR against the whole-genome assemblies of carrot, celery, gotu kola and fennel, with coriander included as a positive control. Simulations were run with ipcress from the exonerate package v. 2.4.0 [26], which reports a product only when both primers of a pair anneal to the same sequence in convergent orientation within a defined product-size window (here 100-1,000 bp). Because no canonical mismatch threshold exists for cross-genus comparison, and preliminary runs showed the result to depend strongly on this parameter, each genome was screened at one, two and three mismatches per primer and the resulting range is reported. A primer pair was scored as transferable to a species when at least one predicted product fell within 50 bp of the size expected in coriander, following the criterion applied to bench amplification by Barboza et al. [27]; products outside that window were treated as non-specific and disregarded, irrespective of their number. For comparison with the homology-based approach often used *in silico*, transferability was also estimated by mapping each primer independently to the target genome and scoring a pair by the lowest of its two hit counts.

***In silico* analysis of SSR polymorphisms between coriander and carrot**

To investigate contractions and expansions of SSRs between the coriander and carrot genomes, and to test whether repeat number is associated with length variation, we compared the number of repeat units of orthologous motifs between the two species. Perfect SSRs identified by MISA (section 2.3) were used; compound repeats were excluded. Motifs were reduced to their canonical form so that equivalent motifs were treated as one. A motif was considered orthologous when present in both genomes. For each orthologous motif, the mean number of repeat units was calculated separately in coriander and carrot, and the allele-length difference (Δal) was computed as the absolute difference in mean repeat number multiplied by the motif length. A motif was scored as expanded or contracted in carrot relative to coriander (taken as the reference state), or as stable when the mean repeat number was identical. To ensure that mean repeat numbers were based on reliable estimates, only motifs supported by at least 10 loci in each species were retained for the length-variation analysis. The relationship between repeat number and Δal was assessed for each SSR type using Spearman's rank correlation, and differences in Δal among SSR types were tested with the Kruskal-Wallis test.

Coriander genetic diversity analysis based on SSRs

A collection of 14 coriander accessions representing diverse geographical origins, including seven Turkish accessions and seven accessions from Afghanistan, the Czech Republic, Denmark, Egypt, Japan, Syria and Russia, was used to evaluate the potential of SSR markers for genetic diversity analysis (Supplementary Table S2). Non-Turkish accessions were obtained from the United States Department of Agriculture (USDA) through the Germplasm Resources Information Network (GRIN; https://www.ars-grin.gov/), with accession numbers indicated in Supplementary Table S2. Turkish accessions correspond to registered varieties and were purchased from local seed markets at their respective provinces and/or cities, indicated in Table S2. Fifteen SSR markers were screened across the collection, and all subsequent diversity analyses were based on the subset that produced polymorphic, scorable bands. For these markers, a matrix was constructed recording the presence or absence of the different alleles at each SSR locus. General statistics, including the number of alleles (na), effective number of alleles (ne) [28], gene diversity (h) [29], Shannon's information index (I) [30] and the polymorphic information content (PIC), were calculated using Popgene v. 1.32 [31]. Analysis of molecular variance (AMOVA) was computed using GenAlEx v. 6.5 [32]. Pairwise genetic similarity values among accessions were estimated according to Jaccard's coefficient, and the resulting matrix was used to construct a dendrogram depicting phenetic relationships among accessions using the unweighted pair-group method with arithmetic means (UPGMA). Principal coordinate analysis (PCoA) was performed to further visualize the distribution of genetic diversity among accessions. The population structure of the 14 coriander accessions was analyzed using a model-based Bayesian clustering approach implemented in STRUCTURE v. 2.1 [33].

Results

Distribution of SSR types across genomic fractions of coriander and related Apiaceae

We recovered 569,531 perfect microsatellites in the 2,116.68 Mbp of assembled coriander genome sequence, which amounts to an overall density of 268.8 SSRs/Mbp (Table 1). Tetranucleotides were, by a wide margin, the most abundant repeat type, making up 39.7% of all SSRs, with trinucleotides (27.0%) and dinucleotides (21.6%) following at comparable levels and the remaining types (penta- at 7.5%, hexa- at 3.4% and heptanucleotides at 0.8%) trailing well behind (Supplementary Table S3). Dinucleotides, however, carried more repeat units on average (12.6) than any other type, so that despite ranking third in frequency they occupied a disproportionate share of the sequence given over to microsatellites.

The same five species were screened on a whole-genome basis for comparison (Supplementary Table S4). Overall SSR density varied about 1.5-fold across the family, from 268.8 SSRs/Mbp in coriander to 390.5 SSRs/Mbp in gotu kola, with carrot (282.9), celery (281.2) and fennel (306.4) in between. Coriander, then, sits at the low-density end of the Apiaceae sampled here. The most abundant type, however, was not conserved across the family: tetranucleotides led in coriander, celery and fennel, whereas dinucleotides were the densest type in carrot and gotu kola (Table 1). The second rank was likewise variable: tetranucleotides ranked second in carrot, trinucleotides second in coriander and gotu kola, and dinucleotides second in celery and fennel (Table 1).

Partitioning the three annotated genomes (coriander, carrot, celery) into genomic fractions revealed a consistent gradient (Fig. 2; Supplementary Table S4). The non-coding fraction tracked the whole genome closely (269.9 SSRs/Mbp in coriander, 297.8 in carrot and 281.6 in celery). Genic sequence was not enriched relative to the genome as a whole (255.2, 238.9 and 238.9 SSRs/Mbp in coriander, carrot and celery), but within the genic compartment

the density was unevenly distributed: introns carried the densest microsatellite content of any fraction we examined (291.9, 303.6 and 303.6 SSRs/Mbp in coriander, carrot and celery, respectively), while exons ran sharply against this trend, holding the lowest densities throughout (197.0, 119.3 and 213.7 SSRs/Mbp). The depletion in exons was accompanied by a shift in composition: trinucleotides, which represent roughly a fifth of SSRs genome-wide, became by far the most frequent type in coding sequence, reaching 44.0% in coriander, 75.0% in carrot and 42.5% in celery. This enrichment for frame-preserving repeats came largely at the expense of tetranucleotides, which fell well below their genome-wide share in all three species; dinucleotides showed a more variable pattern, nearly disappearing in carrot exons (2.8%) yet remaining the second most frequent type in coriander (24.5%) and celery (29.1%).

Distribution of SSR motifs across genomic fractions

Within each repeat type the individual motifs were far from evenly distributed, and their relative ranking shifted predictably as we moved from non-coding into coding sequence (Figs. 3 and 4; Supplementary Tables S5–S9).

Among dinucleotides, AT was the dominant motif in the whole genome of every species, accounting for 63.4% in coriander, 69.3% in celery, 59.9% in carrot, and 60.7% and 63.3% in gotu kola and fennel (Supplementary Table S6). AG ranked second in all five species (25.2% in coriander, 22.8% in carrot, 16.4% in celery, 20.0% in gotu kola and 20.4% in fennel), while CG was negligible everywhere, never exceeding 0.28% of dinucleotide SSRs. The trinucleotide picture was more uniform: AAT was the leading motif in all five species, from 31.7% in gotu kola to 60.9% in fennel (33.7% in coriander, 39.9% in carrot, 43.0% in celery). Among tetranucleotides, the AT-rich AAAT was by far the most frequent motif in every genome (34.9–48.6%), followed by AATT, so that the two together accounted for more than half of tetranucleotide SSRs in each species. Heptanucleotides, the rarest class, were more heterogeneous: AATAATT dominated in coriander (32.2%), AATATAT in carrot (10.2%), AAGTACC in fennel (25.8%), and the homopolymer-adjacent AAAAAAT in gotu kola (14.7%) and celery (4.6%).

The contrast between coding and non-coding sequence was sharpest among the dinucleotides. AT held and slightly strengthened its lead in non-coding DNA (67.1% in coriander, 71.2% in celery, 66.7% in carrot) but collapsed in exons, where it fell below 13% in all three species (12.2%, 10.8%, and 9.5%). In its place, AG became strongly predominant in exonic dinucleotides (67.4% in coriander, 77.4% in celery, and 77.0% in carrot), a near-complete reversal of the genome-wide pattern. Trinucleotides shifted in the same direction: the AT-rich AAT that led non-coding sequence (34.2%, 43.6%, 44.9% in coriander, celery and carrot) fell sharply in exons to under 9% everywhere, giving way to frame-tolerant, often GC-containing motifs such as AAG, ATC and AGC. Tetranucleotides showed the same retreat of AT-rich repeats: AAAT, which made up around half of non-coding tetranucleotide SSRs, dropped to 11–18% in exons, while AAAG and ACAT rose to prominence. The untranslated genic regions fell, as expected, between the two extremes, with AT recovering relative to exons but AG remaining elevated; indeed in introns AAT reached its highest share of any fraction (around 52% in all three species). Taken together, the motif data describe the same coding/non-coding divide seen in the type frequencies: an AT-rich, AT-led landscape across most of the genome, displaced in exons by repeats that read into codons without disrupting them.

SSR marker development and cross-species transferability

Primer pairs were designed for 1,806 coriander SSR loci, with the flanking sequences, expected product sizes, annealing temperatures and GC contents listed in Supplementary Table S11. When these primers were tested in silico with ipcress [26] against the coriander assembly itself,

all 1,806 primer pairs recovered a product of the expected size, confirming that the pipeline behaves as expected on the source genome.

Transferability to the other Apiaceae was assessed in two ways (Table 2). Under the amplicon-based criterion, which counts a pair as transferable only when ipcress predicts a product within 50 bp of the size expected in coriander [27], transferability was low (Supplementary Table S13). Screening each genome at one to three mismatches per primer, carrot recovered 2.2 to 7.6% of the primers, celery 4.3 to 15.6%, fennel 2.8 to 11.0% and gotu kola 0.2 to 4.2%; the full list of transferable primers with their predicted product sizes is given in Supplementary Table S14. The values rose steadily with the mismatch tolerance without reaching a plateau, so the range, rather than any single figure, is the meaningful result. The ranking did not follow phylogenetic distance: celery and fennel, whose assemblies are the largest of the panel, returned more hits than the more closely related carrot, indicating that assembly size and completeness influence recovery alongside true sequence conservation.

By contrast, a homology-only approach in which each primer is mapped independently and a pair scored by its weaker match returned far higher figures for the same primers (Supplementary Table S12): 98.0% in carrot, 91.6% in fennel, 80.1% in gotu kola and 74.9% in celery. The gap between the two approaches, roughly an order of magnitude in carrot, reflects the many primers pairs whose individual primers find a match somewhere in the target genome but do not together define an amplifiable, correctly sized locus. The homology figures are therefore best regarded as an upper bound and the amplicon-based ranges as the more realistic estimate of how these markers behave when transferred across genera.

SSR polymorphisms and length contractions/expansions between coriander and carrot

Length variation between the coriander and carrot genomes was assessed by comparing the number of repeat units of orthologous motifs, reduced to their canonical form, between the two species. Of the motifs recovered, 552 were shared by both genomes and formed the basis of the comparison; coriander carried a further 279 motifs (33.6%) with no counterpart in carrot, while 126 motifs (18.6%) were unique to carrot (Supplementary Table S15). To base each mean repeat number on a reliable sample, only motifs supported by at least ten loci in each species were retained for scoring, leaving 191 motifs for the length-variation analysis (Supplementary Table S16)

Change was nearly universal: of the 191 motifs, 94 (49.2%) revealed expansions and 96 (50.3%) contractions in carrot relative to coriander, while only a single motif (0.5%) was stable (i.e., presented the same size; Fig. 5A). The direction of change varied with repeat type (Supplementary Table S17). Trinucleotides were biased toward expansion (80.0% versus 20.0% contraction) and heptanucleotides likewise (71.4% versus 28.6%), whereas pentanucleotides leaned toward contraction (62.7% versus 35.8%, with 1.5% stable). Di-, tetra- and hexanucleotides were split close to evenly between the two directions, and stable loci were essentially absent from every class.

The magnitude of these changes followed the same logic only loosely (Fig. 5B). Mean allele-length differences (Δal) were largest for the short dinucleotides (3.03 bp) and for the long hexa- and heptanucleotide motifs (1.10 and 2.72 bp), and smallest for the abundant tri-, tetra- and pentanucleotides (0.55, 0.49 and 0.43 bp). Across the informative classes the number of repeat units was positively correlated with Δal (Fig. 5C): Spearman's ρ ranged from 0.53 in tetranucleotides to 0.86 in heptanucleotides (0.66 in tri-, 0.69 in penta- and 0.79 in hexanucleotides), so that loci carrying more repeat units tended to differ more in length between the two genomes. The dinucleotide coefficient ($\rho = 0.80$) rested on only four motifs, was not statistically significant, and was excluded from this range.

SSR-based genetic diversity in coriander germplasm

A panel of 14 coriander accessions of diverse geographical origin was genotyped with 15 SSRs, of which six yielded scorable polymorphic bands (Supplementary Table S18). These six loci resolved 26 alleles, an average of 4.3 per locus and a range of 3 to 6. The markers were informative: polymorphic information content ranged from 0.62 at ssr3 to 0.78 at ssr11, averaging 0.68, with a mean effective allele number of 1.50 and a mean gene diversity of 0.29 (Table 3).

Pairwise genetic similarity (Jaccard) spanned the full available range, from 0.32 to 1.00, with a mean of 0.56 across the panel (Supplementary Table S19). The lowest value separated the Turkish accessions 'Kudret and Gurbuz' and 'Kudret and Kerkuk', while two groups of accessions (i.e., the Japanese accession and the Turkish Kudret, Antakya, and Erbaa; and, conforming the second group, the Gurbuz and Kerkuk) were genetically indistinguishable at these loci (GS = 1.00). The UPGMA dendrogram built from this matrix (cophenetic correlation coefficient was 0.95, indicating a very good representation of the underlying similarity matrix) resolved four principal groups, clustering the accessions with pair-wise genetic similarities above 0.58 (Fig. 6). In the first group, the Turkish Erbaa, Antakya and Kudret clustered with the Japanese accession; a second group clustered two European accessions from Czech Republic and Denmark; a third group included only Turkish materials (Gurbuz, Kerkuk, Arslan, and Pelmus); and the fourth group included three accessions from Syria, Afghanistan, and Russia. Geographic origin and genetic grouping were thus moderately coupled, with Turkish accessions distributed across more than one cluster.

This weak structuring was borne out by both model-based and variance-partitioning analyses. AMOVA assigned the overwhelming majority of variation, 91.6%, to differences within the predefined groups and only 8.4% among them, and the corresponding differentiation was not significant ($\Phi PT = 0.084$, $P = 0.299$; Supplementary Table S20). STRUCTURE, run under a model with $K = 4$, with a mean $LnP(D) = -88.57$ and CLUMPAK [34] similarity of 0.990, recovered two well-defined clusters, one comprising Antakya, Erbaa, Kudret-K, Egypt and Japan and the other Gürbüz, Kerkük, Arslan and Pelmuş, each with membership coefficients above 0.92. A separate set of admixed accessions (Afghanistan, Czech Republic, Denmark, Syria, Russia) had ancestry split almost evenly between the two remaining clusters (Supplementary Table S21). The principal coordinate analysis reproduced this arrangement, separating the two coherent clusters from the admixed group. The concordance of the three methods points to limited overall differentiation among these accessions.

Discussion

At an overall density of 268.8 SSRs/Mbp, coriander sits at the low-density end of the five Apiaceae examined here even though it carries one of the largest genomes in the set. This apparent paradox is consistent with the survey of Palumbo et al. [35], who reported that across Apiaceae the absolute number of SSRs scales with assembly size, with the smallest genome (carrot) returning the fewest repeats and the largest (celery) the most, while the fraction of sequence the repeats occupy shows no such relationship with genome size. Our per-megabase densities behave in exactly that decoupled way, so that a large genome can still be repeat-poor on a density basis. Notably, the two densest assemblies in our comparison are also the two whose draft genomes are least complete, and Palumbo et al. [35] cautioned that short-read assemblies tend to collapse repetitive tracts and can distort repeat counts, so those two figures are best read as provisional. What held firm across every genome was the numerical dominance of tetranucleotides and the concentration of the catalogue in di-, tri-, and tetranucleotide classes, which together accounted for roughly 81% of coriander SSRs, a composition that falls within the 82.6 to 97.6% range that Cavagnaro et al. [8] recorded for these three classes in carrot. This

genome-wide prominence of tetranucleotides departs from the composition recovered whenever the expressed fraction of the family has been surveyed: di- and trinucleotides together made up 97.6% of the fennel EST-SSR catalogue [36], and dinucleotides alone led the transcriptome in celery and parsley [37-39] and in coriander itself [22]. Far from contradicting our result, this pattern follows from it, since a transcriptome survey samples only the genic compartment, where the composition shifts away from the genome-wide tetranucleotide signal in the manner set out below.

Partitioning the annotated genomes into genomic fractions showed that the genic sequence was richer in microsatellites than the genome at large, that the untranslated genic compartment was the densest fraction of all, and that exons were consistently the poorest. The same ordering, with microsatellite density peaking in untranslated regions and falling through coding sequence, was described in maize [40] and in *Arabidopsis* and rice [41], so the gradient we recovered in coriander, carrot and celery appears to be a general feature of plant genomes rather than a family-specific one. The depletion of repeats in exons is expected given that microsatellites mutate rapidly and that expansions or contractions in coding sequence risk disrupting the product, which Vieira et al. [42] invoked to explain the generally lower occurrence of SSRs in plant gene regions. Against that depleted background, trinucleotides rose to become the most frequent or near-most-frequent type in coding sequence. Because the gain or loss of a trinucleotide unit preserves the reading frame whereas an indel that is not a multiple of three shifts it [43, 44], this exonic enrichment of trinucleotides is the predictable signature of selection filtering out frame-disrupting repeats while tolerating those that read cleanly into codons.

A decline in heterologous SSR amplification with phylogenetic distance is a well-established expectation, reported within Apiaceae [8] and beyond it in pear and its Rosaceae relatives [45], in chokecherry across Prunus and other rosaceous genera [46], and in *Carex* [47]; within the family, the cross-genus screen of Sahu et al. [48] found *Centella* primers amplifying 83.3% of loci in the congeneric *Centella javanica* but only about 25% on average across nine other Apiaceae genera, and the grouping of coriander with carrot recovered by Youssef et al. [49] would predict carrot to be the easiest target. Our verified results depart from that expectation on two counts. First, once a correctly sized amplicon is required, the rates are low rather than high (2.2 to 15.6% across species), so the broad conservation implied by a homology screen does not translate into amplification. Second, the ordering did not follow phylogeny at all: the closely related carrot was not the top target, and recovery instead tracked assembly size, with the large celery and fennel assemblies scoring highest. The phylogenetic-decay pattern seen in wet-lab studies may still hold in the laboratory, but it is not what an amplicon-based in silico screen recovers across these particular genome assemblies.

The motif data told the same story at finer resolution. AT was the leading dinucleotide in the whole genome of every species and CG was almost absent, a pattern reported for Apiaceae [35] and that Qin et al. [50] found to strengthen with repeat number in dicotyledonous plants generally, where AT/TA rises and CG/GC declines as arrays lengthen. The trinucleotide picture was led by AAT in coriander, in agreement with the AAT/ATT preference recorded for the family [35]. The sharpest contrast was again between coding and non-coding sequence; the genome-wide dominance of AT among dinucleotides collapsed in exons, where AG took its place, and the AT-rich tetranucleotides gave way in coding sequence to G-containing motifs such as AAAG and ACAT. The same exonic shift toward AG/CT repeats is recovered wherever the transcribed fraction of coriander is sampled [22] and wherever genomic and transcribed carrot SSRs are compared [8]. The fact that this enrichment recurs across the expressed sequence of other Apiaceae, including fennel, celery, parsley, and Centella asiatica [37, 38, 48, 51], indicates that within-codon selection, rather than regional base composition alone, governs which dinucleotides persist in coding DNA. The exonic shift toward AG-type dinucleotides

and G-containing tetranucleotides therefore reflects the same constraint seen in the type frequencies, namely repeats that can be carried inside codons without disrupting them.

The verified transferability we recorded is far lower than a homology-only screen of the same primers would suggest, and it sits much closer to the levels seen in the laboratory when Apiaceae primers are moved across genera, such as the roughly 30% amplification in cilantro reported by Cavagnaro et al. [8], the 23% transfer of carrot SSRs into fennel found by Aiello et al. [52], who attributed the shortfall to poor conservation of the flanking regions and a high mutation rate among distantly related taxa, and the 39% of carrot loci that Choudhary et al. [53] recovered in coriander, of which only 5% proved polymorphic. The contrast between our two estimates is itself informative: a homology screen counts every conserved priming site and returned 75 to 98% here, whereas requiring both primers to define a correctly sized amplicon collapses the figures by roughly an order of magnitude. This gap quantifies, on a single primer set and genome panel, how far the homology shortcut overstates cross-genus transferability, and it cautions against reading *in silico* homology counts as amplification rates. As Barboza et al. [27] applied it at the bench, a transfer was credited only when a product of the expected size appeared; imposing the same size criterion computationally is what brings our *in silico* estimates into line with wet-lab experience. That recovery tracked assembly size as much as phylogenetic proximity, with the larger celery and fennel assemblies outscoring the more closely related carrot, reinforces this reading, since a more complete assembly offers more opportunities for a size-matched product to be found. It remains important, as Cavagnaro et al. [8] emphasised, that recovery of a product does not by itself prove locus homology, since size homoplasy can converge non-homologous fragments, so the markers flagged here as transferable are best treated as candidates pending bench confirmation. With that caveat, the breadth of conservation is notable for an annual herb, given that faster generation turnover in herbaceous taxa is expected to accelerate genomic differentiation relative to woody perennials, as Butler et al. [54] argued; the 1,806 primers developed here still represent a substantial genomic resource for coriander and a starting point for cross-genus work in Apiaceae, helping to close the marker gap that Palumbo et al. [35] identified for the family outside carrot.

Comparing repeat number at orthologous motifs between the two genomes let us score length variation at 552 shared motifs, of which 191 had sufficient locus support for analysis. A third of the coriander motifs (33.6%) had no counterpart in carrot, while a smaller fraction of carrot motifs (18.6%) were absent from coriander. A substantial fraction of species-specific loci is unsurprising given that structural rearrangements, copy-number changes and indels routinely generate and erase repeat-bearing segments between genomes, as reviewed by Dolatabadian and Fernando [55] and documented for spontaneous restructuring in Brassica [56]. Within Apiaceae itself, the assembled coriander genome revealed extensive and unbalanced gene loss between coriander and carrot following the two tetraploidizations the species share, with most orthologous positions retaining fewer copies than expected from the ancestral state [7], a lineage-specific fractionation that offers a direct, family-internal source for the loci we found confined to one genome. Among the shared loci, polymorphism depended strongly on repeat type. Change was nearly universal across the informative motifs, with expansions and contractions roughly balanced and stable loci essentially absent, while the direction of change varied by type, trinucleotides and heptanucleotides leaning toward expansion and pentanucleotides toward contraction. The number of repeat units was positively correlated with the allele-length difference in every class, with Spearman coefficients ranging from 0.53 to 0.86 across the informative types, so that longer arrays within each type expanded or contracted more between the two genomes. This is the expectation set by Vieira et al. [42], who noted that longer and purer repeats mutate more often through the addition or loss of whole units, and the mechanism behind those small gains and losses is the slipped-strand mispairing described earlier [43, 44], in which the replication complex reanneals one or two units out of

register and so drives the stepwise expansions and contractions that dominated our data. The occasional large jumps at some long-motif loci sit outside the range of simple slippage and are better explained by larger indels or repeat-mediated rearrangements, a process whose dependence on repeat abundance has been documented in other plant systems, such as the *Medicago* plastomes [57], where the degree of rearrangement rose with the abundance of repetitive sequence. It is also possible that some of these large length variations could arise from InDels outside the microsatellite sequences. Overall, the practical reading of these patterns is straightforward: repeat-rich loci are the most rewarding to target when polymorphic markers are the goal.

Applied to a panel of 14 coriander accessions, six polymorphic loci resolved 26 alleles, an average of 4.3 per locus, with a mean PIC of 0.68. As a measure of marker performance, this allelic richness is close to the 4.47 alleles per locus reported for genomic SSRs in *Viburnum* [58] and the 5.2 found with EST-SSRs in *Angelica dahurica* [36], while the informativeness exceeds the mean PIC of 0.267 obtained with carrot-derived SSRs in fennel [52] and falls within the 0.264 to 0.728 range recorded in Chinese cherry [59]. The closest benchmark, however, comes from coriander itself: genotyping 14 Indian accessions with transcriptome-derived SSRs, Tulsani et al. [22] recovered a comparable richness of roughly four alleles per locus alongside similarly modest diversity, and the low yield of polymorphic carrot loci transferred into coriander by Choudhary et al. [53] points the same way. The overall level of diversity is nonetheless modest when set against the 19 alleles per locus and expected heterozygosity of 0.84 that Cavagnaro et al. [8] measured across a broad carrot collection, a contrast that reflects both their larger and more diverse accession set and the narrow genetic base of cultivated coriander. That narrow base is borne out by the population analyses: AMOVA placed 91.6% of the variation within groups and only 8.4% among them, and the differentiation was not significant, a weak structuring consistent with the cultivated origin of the material and comparable to the gap between observed and expected heterozygosities traced to the same cause [58]. Even so, the markers were polymorphic enough to separate the accessions, in keeping with the high information content that makes SSRs more discriminating than biallelic markers [42].

Conclusions

This study provides a genome-wide catalogue of microsatellites for coriander and sets it within a comparative framework of four related Apiaceae. Coriander proved to lie at the low-density end of the family, yet it shares with the other species a set of broadly conserved features: a landscape dominated by tetra- or dinucleotides depending on the species, an AT-rich motif composition that shifts toward frame-preserving, G-containing repeats in coding sequence, and a marked concentration of microsatellites in introns and other untranslated genic regions, a combination that appears general to plant genomes rather than specific to the family.

From this catalogue we designed 1,806 primer pairs and assessed their cross-species transfer in silico by two approaches. When transfer was scored realistically, requiring a correctly sized amplicon, verified transferability was low across the family (below 16% in every species) and, unexpectedly, did not track phylogenetic distance: recovery was governed more by the size and completeness of each target assembly than by relatedness to coriander, so that the closest species, carrot, was not the most accessible. A homology-only screen returned far higher figures, but the roughly ten-fold gap between the two shows that homology counts conserved priming sites rather than amplifiable loci and therefore overstates true transferability. These primers nonetheless constitute the first in silico-characterised set of cross-genus SSR markers for Apiaceae and offer a starting point for comparative work that has

been constrained in the family beyond carrot, provided the candidates are confirmed at the bench. The comparison between coriander and carrot further showed that length variation at orthologous loci is governed chiefly by repeat number, with longer arrays diverging more and short-motif repeats proving the most labile; expansions and contractions were nearly balanced and stable loci essentially absent, a pattern consistent with replication slippage that identifies repeat-rich loci as the most productive targets for polymorphic markers. When applied to a panel of cultivated accessions, six of these markers resolved informative but modest diversity with little population structure, in agreement with the narrow genetic base documented for the crop.

Bench validation of the most transferable primers and their application to a broader and wilder set of accessions would convert this resource into a working toolkit for the genetic improvement and conservation of coriander and its Apiaceae relatives.

Author contributions: MY and PFC planned and conceptualized the research. CM and KB conducted bioinformatic analyses and drafted the original manuscript. AM conducted bioinformatic analysis. GS, MK, and MAP conducted field experiment and wet lab analysis of SSR-based genetic diversity. CA, MY, and PFC obtained funding and supervised. All the authors revised and agreed on the final version of the manuscript.

Funding: This study was partially funded by the 'Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación', Argentina, through grant PICT-2020-SERIE-A-3982.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors have no competing interests

Data availability: All the data used in this study was provided in Supplementary Tables S1-S22.

References

1. Seidemann J. World spice plants. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005.
2. Laribi B, Kouki K, M'Hamdi M, Bettaieb T. Coriander (*Coriandrum sativum* L.) and its bioactive constituents. Fitoterapia. 2015;103:9–26. https://doi.org/10.1016/j.fitote.2015.03.012
3. Sachan AKR, Kumar S, Kumari K, Singh D. Medicinal uses of spices used in our traditional culture: World wide. J Med Plants Stud. 2018;6:116–22.
4. Arora V, Adler C, Tepikin A, Ziv G, Kahane T, Abu-Nassar J, et al. Wild coriander: an untapped genetic resource for future coriander breeding. Euphytica. 2021;217:138. https://doi.org/10.1007/s10681-021-02870-4

5. Hedge IC, Lamond JM. *Coriandrum* L. In: Davis P, editor. Flora of Turkey and the East Aegean Islands. Vol 4. Edinburgh: University of Edinburgh Press; 1972. pp. 320–31.
6. Song X, Wang J, Li N, Yu J, Meng F, Wei C, et al. Deciphering the high-quality genome sequence of coriander that causes controversial feelings. Plant Biotechnol J. 2020;18(6):1444–56. https://doi.org/10.1111/pbi.13310
7. Kadereit JW, Bittrich V. Flowering plants. Eudicots: Apiales, Gentianales (except Rubiaceae). The families and genera of vascular plants. Vol 15. Cham: Springer; 2018. https://doi.org/10.1007/978-3-319-93605-5
8. Cavagnaro PF, Chung SM, Manin S, Yildiz M, Ali A, Alessandro MS, et al. Microsatellite isolation and marker development in carrot: genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across Apiaceae. BMC Genomics. 2011;12:386. https://doi.org/10.1186/1471-2164-12-386
9. Sayed-Ahmad B, Talou T, Saad Z, Hijazi A, Merah O. The Apiaceae: Ethnomedicinal family as source for industrial uses. Ind Crops Prod. 2017;109:661–71. https://doi.org/10.1016/j.indcrop.2017.09.027
10. Aćimović MG. Nutraceutical potential of Apiaceae. In: Mérillon JM, Ramawat K, editors. Bioactive molecules in food. Cham: Springer; 2019. https://doi.org/10.1007/978-3-319-78030-6_17
11. Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, et al. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics. 2010;11:569. https://doi.org/10.1186/1471-2164-11-569
12. Moges AD, Admassu B, Belew D, Yesuf M, Njuguna J, Kyalo M, et al. Development of microsatellite markers and analysis of genetic diversity and population structure of *Colletotrichum gloeosporioides* from Ethiopia. PLoS One. 2016;11(3):e0151257. https://doi.org/10.1371/journal.pone.0151257
13. Xu X, Wang B, Zhou C, Yang H, Zhong X, Li W, et al. 2023. Characterization of highly polymorphic microsatellite markers for the chinese monal (*Lophophorus lhuysii*, Galliformes) using Illumina MiSeq sequencing. Mol Biol Rep. 2023;50:3903–8. https://doi.org/10.1007/s11033-022-08151-0
14. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1996;1(7):215–22. https://doi.org/10.1016/1360-1385(96)86898-1
15. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics. 2003;270:315–23. https://doi.org/10.1007/s00438-003-0921-4
16. Gupta PK, Varshney R. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 2000;113:163–85. https://doi.org/10.1023/A:1003910819967.
17. Kanzana G, Musaza J, Wu F, Ouyang Z, Wang Y, Ma T, et al. Genome-wide development and application of miRNA-SSR markers in Melilotus genus. Physiol Mol Biol Plants. 2021;27(10):2269–82. https://doi.org/10.1007/s12298-021-01086-z
18. Choudhary S, Naika MBN, Meena RD. Development and characterization of genic SSR-FDM for stem gall disease resistance in coriander (*Coriandrum sativum* L.) and its cross species transferability. Mol Biol Rep. 2021;48:3963–70. https://doi.org/10.1007/s11033-021-06396-9
19. Šarhanová P, Pfanzelt S, Brandt R, Himmelbach A, Blattner FR. SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. Ecol Evol. 2018;8(22):10817–33. https://doi.org/10.1002/ece3.4533

20. Wang Y, Lan Q, Zhao X, Wang L, Yu W, Wang B, et al. The complete mitochondrial genome of *Coriandrum sativum*. Mitochondrial DNA B Resour. 2021;6(8):2391–2. https://doi.org/10.1080/23802359.2021.1951131
21. Song X, Nie F, Chen W, Ma X, Gong K, Yang Q, et al. Coriander Genomics Database: a genomic, transcriptomic, and metabolic database for coriander. Hortic Res. 2020;7:55. https://doi.org/10.1038/s41438-020-0261-0
22. Tulsani NJ, Hamid R, Jacob F, Umretiya NG, Nandha AK, Tomar RS, et al. Transcriptome landscaping for gene mining and SSR marker development in Coriander (*Coriandrum sativum* L.). Genomics. 2020;112(2):1545–53. https://doi.org/10.1016/j.ygeno.2019.09.004
23. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 1989;17:6463–70.
24. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003;106:411–22.
25. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols. Methods Mol Biol. Vol 132. Totowa, NJ: Humana Press; 2000;132:365–86. https://doi.org/10.1385/1-59259-192-2:365
26. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinfor. 2005;6:31. https://doi.org/10.1186/1471-2105-6-31
27. Barboza K, Beretta V, Kozub PC, Salinas C, Morgenfeld MM, Galmarini CR, Cavagnaro PF. Microsatellite analysis and marker development in garlic: distribution in EST sequence, genetic diversity analysis, and marker transferability across Alliaceae. Mol Genet Genomics. 2018;293(5):1091-106. https://doi.org/10.1007/s00438-018-1442-5
28. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;49(4):725–38. https://doi.org/10.1093/genetics/49.4.725
29. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA. 1973;70(12):3321–23. https://doi.org/10.1073/pnas.70.12.3321
30. Lewontin RC. The apportionment of human diversity. In: Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary biology. Vol 6. New York: Springer; 1972. pp. 381–98. https://doi.org/10.1007/978-1-4684-9063-3_14
31. Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX. POPGENE, version 1.32: the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Canada; 1999.
32. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. Bioinformatics. 2012;28(19):2537–9. https://doi.org/10.1093/bioinformatics/bts460
33. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59. https://doi.org/10.1093/genetics/155.2.945
34. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Res. 2015;15(5):1179–91. https://doi.org/10.1111/1755-0998.12387
35. Palumbo F, Vannozzi A, Barcaccia G. Impact of genomic and transcriptomic resources on Apiaceae crop breeding strategies. Int J Mol Sci. 2021;22(18):9713. https://doi.org/10.3390/ijms22189713
36. Chen C, Chen Y, Huang W, Jiang Y, Zhang H, Wu W. Mining of simple sequence repeats (SSRs) loci and development of novel transferability-across EST-SSR markers from de novo transcriptome assembly of *Angelica dahurica*. PLoS ONE. 2019;14(8):e0221040. https://doi.org/10.1371/journal.pone.0221040

37. Li M-Y, Wang F, Jiang Q, Ma J, Xiong A-S. Identification of SSRs and differentially expressed genes in two cultivars of celery (*Apium graveolens* L.) by deep transcriptome sequencing. Hortic Res. 2014;1:10. https://doi.org/10.1038/hortres.2014.10
38. Li M-Y, Tan H-W, Wang F, Jiang Q, Xu Z-S, Tian C, et al. De novo transcriptome sequence assembly and identification of AP2/ERF transcription factor related to abiotic stress in parsley (*Petroselinum crispum*). PLoS ONE 2014;9(9):e108977. https://doi.org/10.1371/journal.pone.0108977
39. Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA. 1994;91:5466-70.
40. Zhao M, Shu G, Hu Y, Cao G, Wang Y. Pattern and variation in simple sequence repeat (SSR) at different genomic regions and its implications to maize evolution and breeding. BMC Genom. 2023;24:136. https://doi.org/10.1186/s12864-023-09156-0
41. Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. Genome Biol. 2006;7(2):R14. https://doi.org/10.1186/gb-2006-7-2-r14
42. Vieira ML, Santini L, Diniz AL, Munhoz CF. Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol. 2016;39(3):312–28. https://doi.org/10.1590/1678-4685-GMB-2016-0027
43. Sehn JK. Insertions and deletions (indels). In: Elsevier eBooks; 2014. pp. 129–50. https://doi.org/10.1016/b978-0-12-404748-8.00009-5
44. Low D, Van Der Woude M. Phase variation. In: Elsevier eBooks; 2013. pp. 286–9. https://doi.org/10.1016/b978-0-12-374984-0.01144-x
45. Fan L, Zhang MY, Liu QZ, Li LT, Song Y, Wang LF. Transferability of newly developed pear SSR markers to other Rosaceae species. Plant Mol Biol Rep. 2013;31(6):1271–82. https://doi.org/10.1007/s11105-013-0586-z
46. Wang H, Walla JA, Zhong S, Huang D, Dai W. Development and cross-species/genera transferability of microsatellite markers discovered using 454 genome sequencing in chokecherry (*Prunus virginiana* L.). Plant Cell Rep. 2012;31(11):2047–55. https://doi.org/10.1007/s00299-012-1315-z
47. Șuteu D, Pușcaș M, Băcilă I, Balázs ZR, Choler P. Cross-species transferability of specific SSR markers from *Carex curvula* (Cyperaceae) to other Carex species. Diversity. 2024;16(2):73. https://doi.org/10.3390/d16020073
48. Sahu J, Das Talukdar A, Devi K, Dutta Choudhury M, Barooah M, Modi MK, et al. E-microsatellite markers for *Centella asiatica* (Gotu Kola) genome: validation and cross-transferability in Apiaceae family for plant omics research and development. OMICS J Integr Biol. 2015;19(1):52–65. https://doi.org/10.1089/omi.2014.0113
49. Youssef D, El-Bakatoushi R, Elframawy A, El-Sadek L, Badan GE. Molecular phylogenetic study of flavonoids in medicinal plants: a case study family Apiaceae. J Plant Res. 2023;136(3):305–22. https://doi.org/10.1007/s10265-023-01442-y
50. Qin Z, Wang Y, Wang Q, Li A, Hou F, Zhang L. Evolution analysis of simple sequence repeats in plant genome. PLoS ONE 2015;10(12):e0144108. https://doi.org/10.1371/journal.pone.0144108
51. Palumbo F, Vannozzi A, Vitulo N, Lucchin M, Barcaccia G. The leaf transcriptome of fennel (*Foeniculum vulgare* Mill.) enables characterization of the t-anethole pathway and the discovery of microsatellites and single-nucleotide variants. Sci Rep. 2018;8:10459. https://doi.org/10.1038/s41598-018-28775-2
52. Aiello D, Ferradini N, Torelli L, Volpi C, Lambalk J, Russi L, et al. Evaluation of cross-species transferability of SSR markers in *Foeniculum vulgare*. Plants. 2020;9(2):175. https://doi.org/10.3390/plants9020175

53. Choudhary S, Jethra G, Sharma R, Verma AK. Microsatellite in coriander: a cross species amplification within Apiaceae family. Int J Curr Microbiol Appl Sci. 2017;6(5):2714–21. https://doi.org/10.20546/ijcmas.2017.605.304
54. Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, et al. Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. BMC Genom. 2017;18(1):397. https://doi.org/10.1186/s12864-017-3782-7
55. Dolatabadian A, Fernando WGD. Genomic variations and mutational events associated with plant–pathogen interactions. Biology. 2022;11(3):421. https://doi.org/10.3390/biology11030421
56. Orantes-Bonilla M, Makhoul M, Lee H, Chawla HS, Vollrath P, Langstroff A, et al. Frequent spontaneous structural rearrangements promote rapid genome diversification in a *Brassica napus* F1 generation. Front Plant Sci. 2022;13:1057953. https://doi.org/10.3389/fpls.2022.1057953
57. Wu S, Chen J, Li Y, Liu A, Li A, Yin M, et al. (2021). Extensive genomic rearrangements mediated by repetitive sequences in plastomes of Medicago and its relatives. BMC Plant Biol. 21:421. https://doi.org/10.1186/s12870-021-03202-3
58. Hamm TP, Nowicki M, Boggess SL, Klingeman WE, Hadziabdic D, Huff ML, et al. Development and characterization of 15 novel genomic SSRs for *Viburnum farreri*. Plants. 2021;10(3):487. https://doi.org/10.3390/plants10030487
59. Liu Z, Zhang J, Wang Y, Wang H, Wang L, Zhang L, et al. Development and cross-species transferability of novel genomic-SSR markers and their utility in hybrid identification and trait association analysis in Chinese cherry. Horticulturae 2022;8(3):222. https://doi.org/10.3390/horticulturae8030222

Table 1. Distribution of SSRs in genomic sequences of coriander and related Apiaceae species.

SSR Type	*Coriandrum sativum*				Microsatellite density in related species (SSR/Mbp)				All species mean
	Count	Rel. freq. (%)	Mean repeat number	Density (SSR/Mbp)	*D. carota*	*A. graveolens*	*C. asiatica*	*F. vulgare*	
Genomic									
Dinucleotide	122,917	21.6	12.6	58.0	91.2	83.8	124.4	78.9	87.3
Trinucleotide	153,984	27.1	4.9	72.7	61.5	67.6	112.3	75.6	77.9
Tetranucleotide	226,108	39.7	3.1	106.7	78.6	102.5	104.1	93.1	97.0
Pentanucleotide	42,919	7.5	3.1	20.3	35.2	18.5	33.4	42.4	30.0
Hexanucleotide	19,088	3.4	3.1	9.0	13.8	7.5	13.3	12.7	11.3
Heptanucleotide	4,515	0.8	3.2	2.1	2.7	1.5	3.0	3.8	2.6
Total/mean	**569,531**	**100.0**	**5.7**	**268.8**	**283.0**	**281.4**	**390.5**	**306.5**	**306.0**
Total seq. (Mbp)				2,118.7	421.5	3,332.6	455.6	1,011.4	
GC content (%)				34.8	34.8	35.7	34.4	33.6	
Non-coding sequences									
Dinucleotide	109,455	20.6		55.5	94.1	82.8			77.5
Trinucleotide	145,514	27.4		73.8	62.0	68.0			67.9
Tetranucleotide	214,917	40.4		109.0	83.8	103.6			98.8
Total/mean	**469,886**	**88.3**		**238.3**	**239.9**	**254.4**			**244.2**
Total seq. (Mbp)				1,971.9	313.3	3,200.7			
Genic sequences									
Dinucleotide	13,621	36.6		92.8	82.7	107.3			94.3
Trinucleotide	8,447	22.5		57.5	59.9	57.7			58.4
Tetranucleotide	11,107	29.6		75.7	63.2	74.2			71.0
Total/mean	**33,175**	**88.5**		**226.0**	**205.8**	**239.2**			**223.6**
Total seq. (Mbp)				146.8	108.2	131.9			
Untranslated genic regions									
Dinucleotide	10,877	41.6		121.4	126.2	130.2			125.9
Trinucleotide	3,476	13.3		38.8	43.4	40.8			41.0
Tetranucleotide	8,732	33.4		97.4	89.7	91.7			93.0
Total/mean	**23,085**	**88.3**		**257.6**	**259.3**	**262.7**			**259.9**
Total seq. (Mbp)				89.6	69.9	87.6			
Exons									
Dinucleotide	2,755	24.5		48.2	3.3	62.1			37.9
Trinucleotide	4,955	44,0		86.6	89.5	90.9			89.0
Tetranucleotide	2,336	20.7		40.8	14.1	38.9			31.3
Total/mean	**10,046**	**89.2**		**175.6**	**106.8**	**191.9**			**158.1**
Total seq. (Mbp)				57.2	38.3	44.3			

Table 2. Cross-species transferability of the 1,806 coriander SSR primer pairs, estimated *in silico* by two approaches.

Species	Genome (Mbp)	Homology-based (%)	Amplicon-based (%)
Coriandrum sativum	2,118.7	100	100
Daucus carota	421.5	98.0	2.2 – 7.6
Apium graveolens	3,332.6	74.9	4.3 – 15.6
Centella asiatica	455.6	80.1	0.2 – 4.2
Foeniculum vulgare	1,011.4	91.6	2.8 – 11.0

Table 3. Genetic diversity statistics for six polymorphic SSR loci in 14 coriander accessions.

Locus	Alleles	Na	Ne	He	I	PIC
ssr1	3	1.667	1.547	0.300	0.428	0.643
ssr4	5	1.800	1.435	0.259	0.394	0.687
ssr5	4	1.75	1.636	0.344	0.488	0.652
ssr6	6	1.833	1.525	0.308	0.459	0.713
ssr11	5	1.800	1.259	0.196	0.328	0.784
ssr3	3	1.667	1.577	0.308	0.436	0.624
Mean	4.3	1.753	1.497	0.286	0.422	0.684
SD	1.2	0.072	0.134	0.052	0.056	0.057

Na: observed number of alleles; Ne: effective number of alleles; He: Nei's expected heterozygosity; I: Shannon's information index; PIC: polymorphism information content.

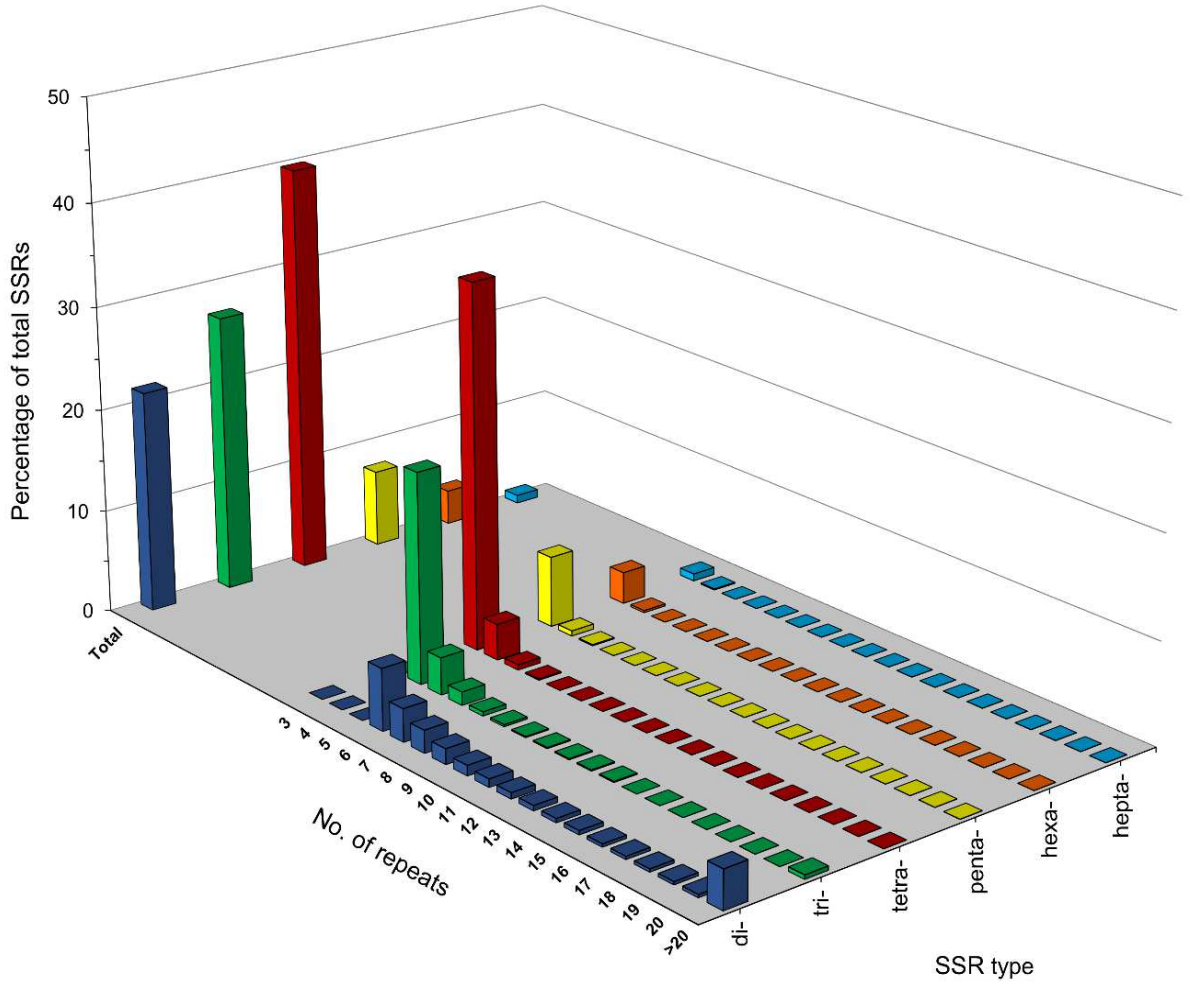


Fig. 1 Relative frequency (%) of the six SSR types identified in the coriander genome. Values were calculated from a total of 569,531 SSRs detected in 2,119 Mbp of genomic DNA sequence of the reference coriander genome [6]. Detailed information for this figure is presented in Supplementary Table S3.

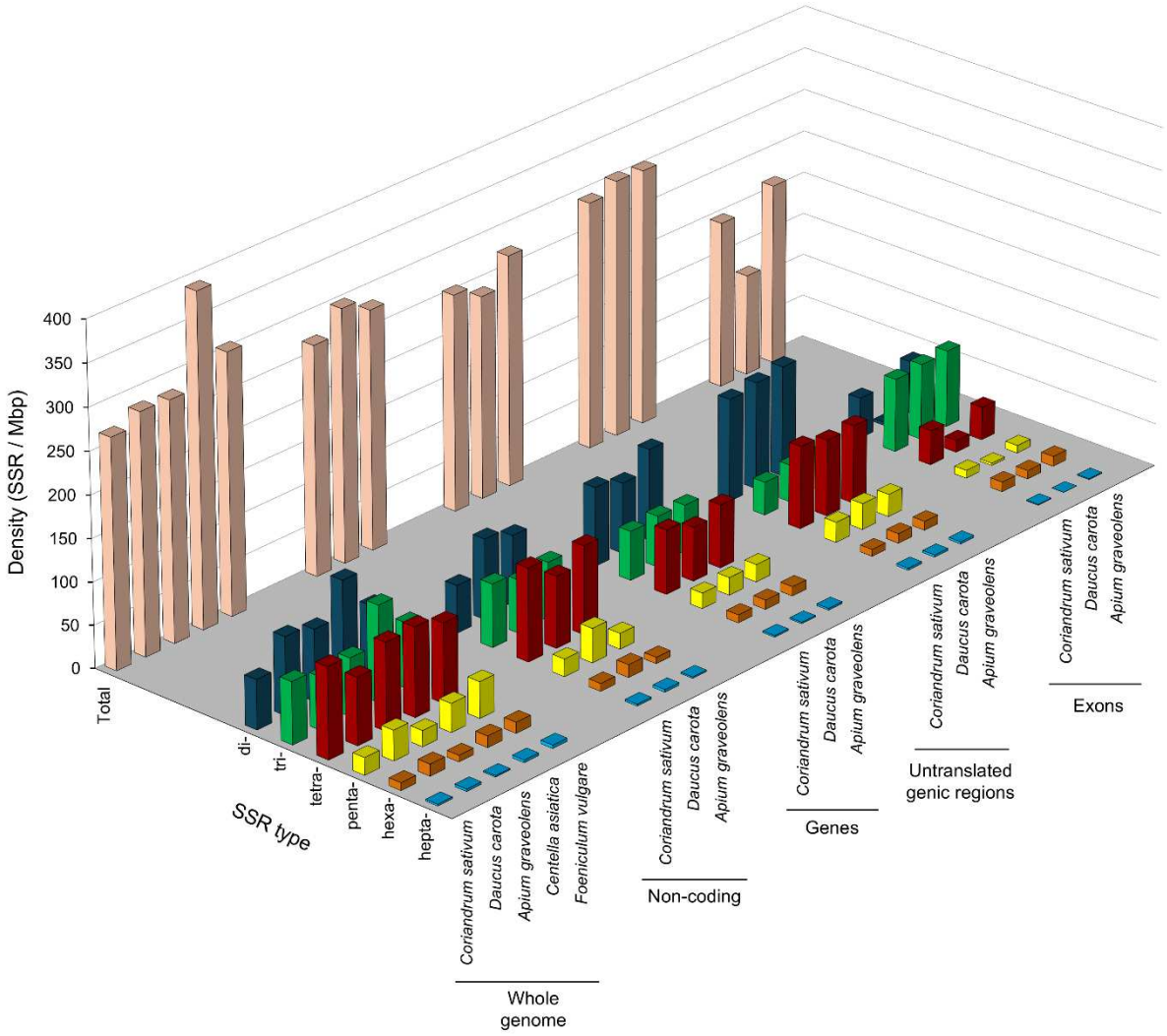


Fig. 2 Density (SSRs/Mbp) of the different SSR types in the whole-genome sequences of coriander, carrot, celery, gotu kola and fennel, and in the non-coding, genic, untranslated genic and exonic fractions of the three annotated genomes (coriander, carrot and celery). Detailed information for this figure is presented in Supplementary Table S4.

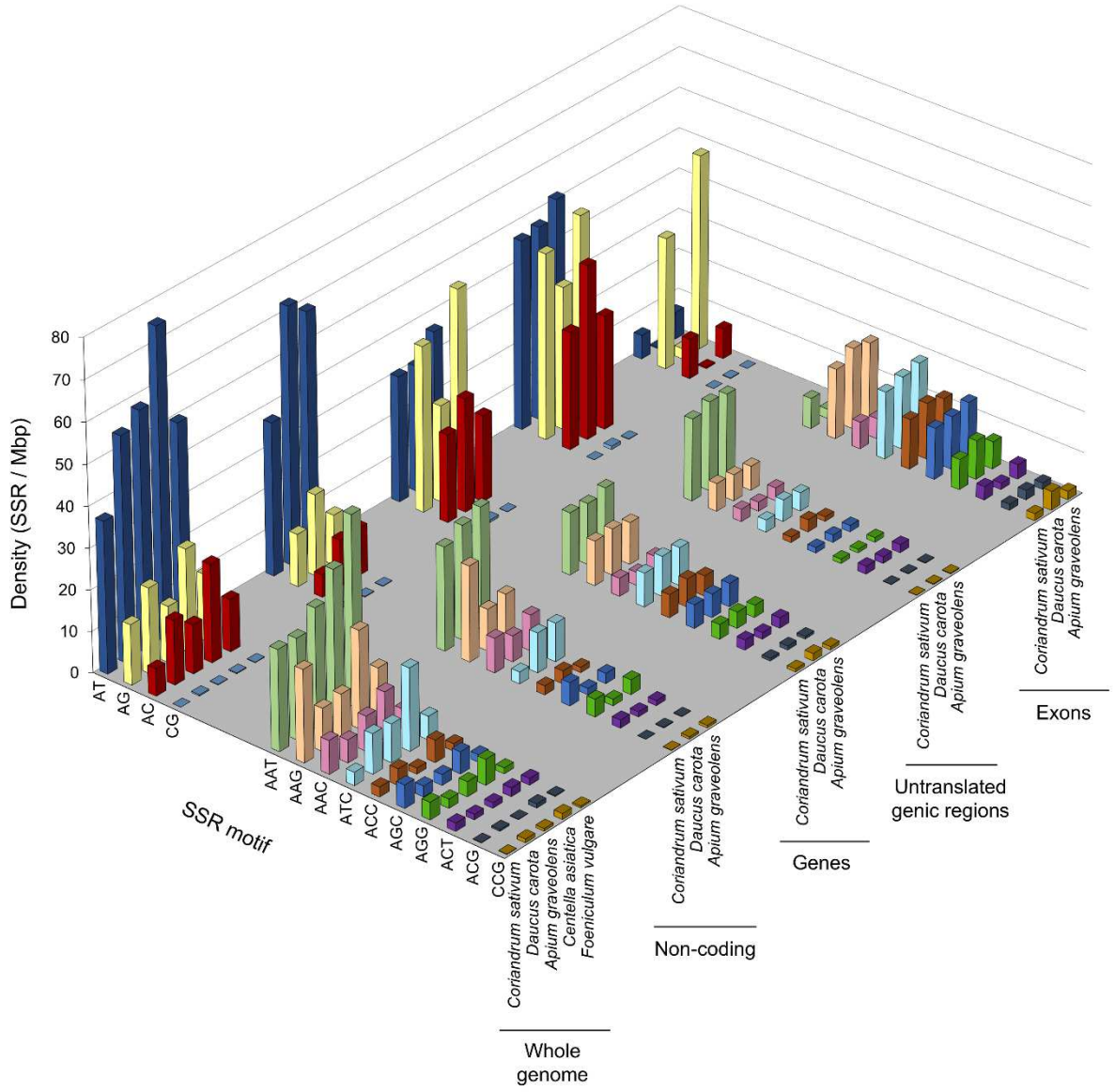


Fig. 3 Distribution of different SSR motifs of di- and trinucleotide repeats in whole genome sequences of the Apiaceae coriander, carrot, celery, gotu kola, and fennel; and in specific genomic fractions of non-coding, genes, untranslated genic regions, and exons. Frequency values were expressed as number of repeats per million base pairs of sequence. Detailed information of the data presented in this figure is presented in Supplementary Tables S5-S9.

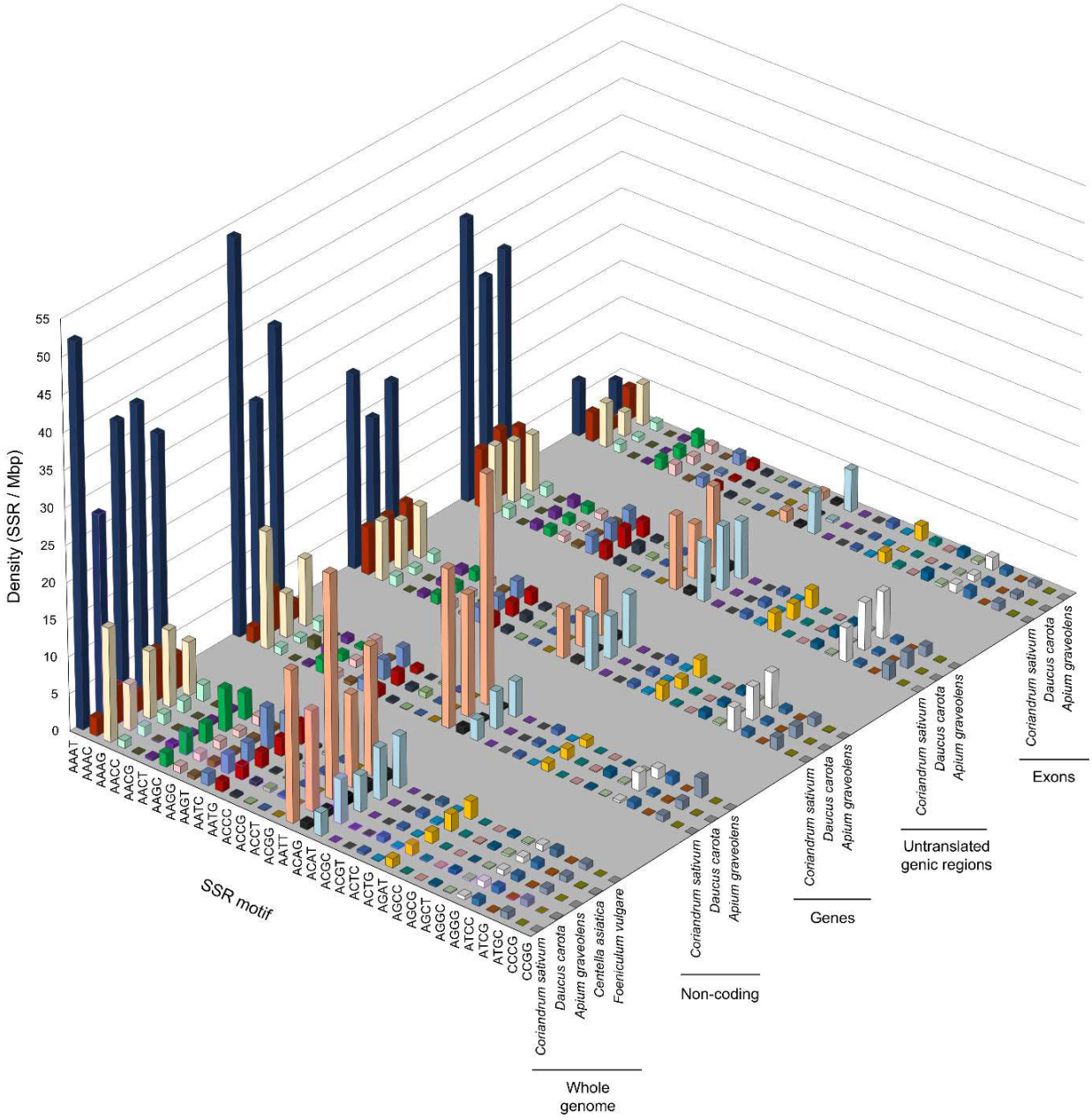


Fig. 4 Distribution of different SSR motifs of tetranucleotide repeats in whole genome sequences of the Apiaceae coriander, carrot, celery, gotu kola, and fennel; and in specific genomic fractions of non-coding, genes, untranslated genic regions, and exons. Frequency values were expressed as the number of repeats per million base pairs of sequence. Detailed information of the data presented in this figure is presented in Supplementary Tables S5-S9.

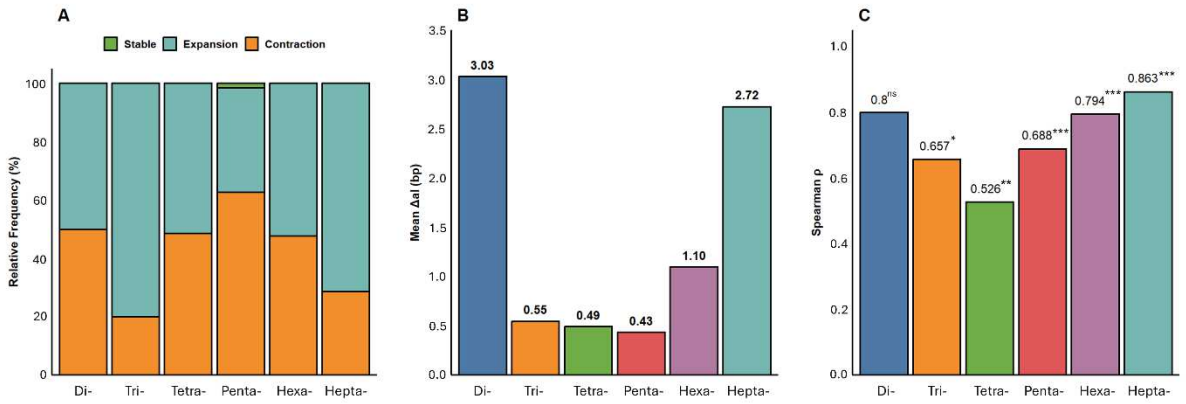


Fig. 5 Length variation at orthologous SSR loci between the coriander and carrot genomes, by repeat type. (A) Relative frequency (%) of loci scored as stable, expanded or contracted in carrot relative to coriander, the latter taken as the reference (initial) state of the repeat. (B) Mean allele-length difference (Δal, in bp) between the two genomes. (C) Spearman's rank correlation coefficient (ρ) between the number of repeat units and Δal for each SSR type. Significant correlations are indicated with asterisks, for $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); ns. non-significant.

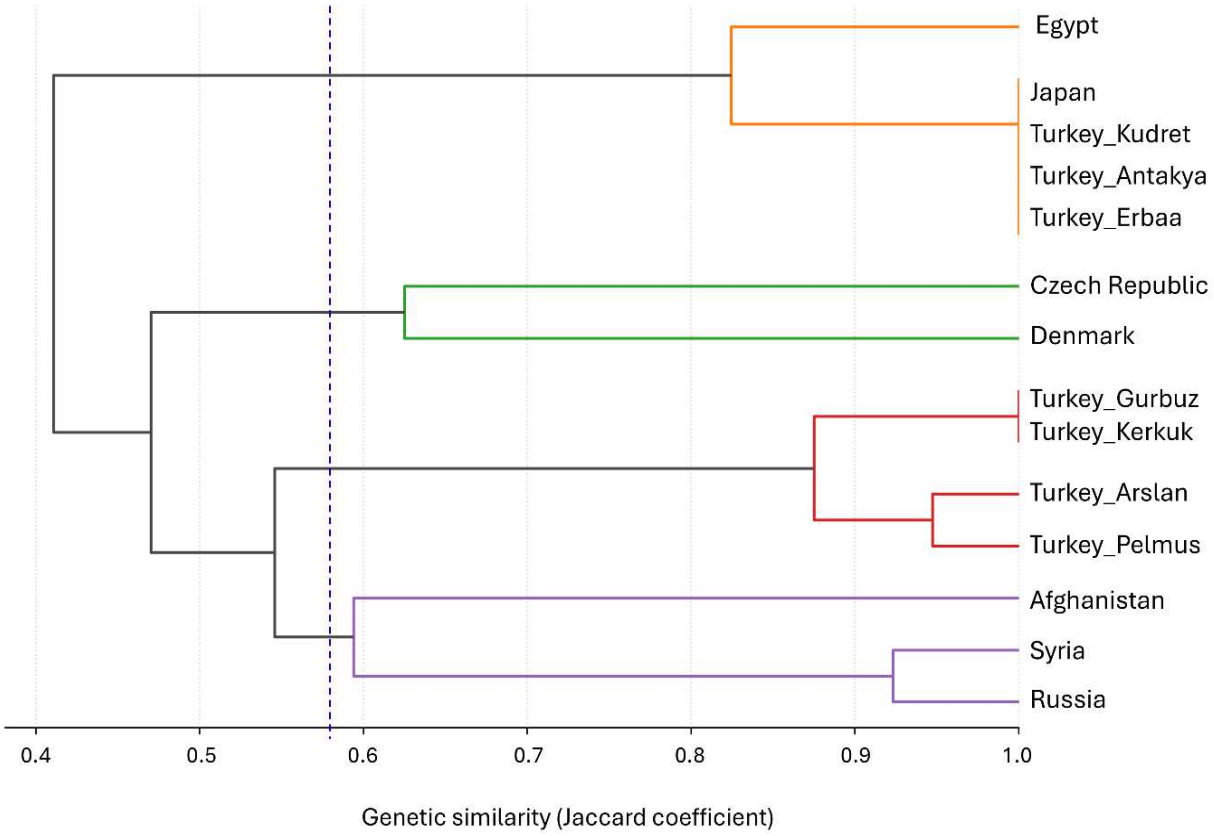


Fig. 6 UPGMA dendrogram based on six polymorphic SSR loci showing genetic relationships among 14 coriander accessions. Colored clusters include accessions with pair-wise genetic similarities > 0.58 (denoted by the blue dashed vertical line). The accessions are labeled based on their country of origin and, for Turkish materials, their specific location.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1TablesS1S21revised.xlsx