Apiaceae Crop Breeding Strategies

Subjects: Plant Sciences
Contributor: Fabio Palumbo

The economic influence of a species is a pivotal force driving the development of new relevant knowledge. Indeed, the information available for economically relevant species is expected to be far more abundant than the data available for species of little interest. To this end, the present review aims to highlight both the lack and progress achieved in the genomics and transcriptomics of this family, with a particular emphasis on their applications for breeding purposes.

Keywords: marker-assisted selection ; marker-assisted breeding ; SSR markers ; genome assembly ; reproductive barriers ; cytoplasmic male-sterility ; nuclear male-sterility

1. General Introduction

The Apiaceae family (previously known as Umbelliferae) is a nearly cosmopolitan taxon that includes about 3820 species organized into 466 genera. According to the most recent and comprehensive phylogenetic study based on the analysis of the plastid *rpl16* intron and *trnD-trnY-trnE-trnT* intergenic spacers in 144 genera, the Apiaceae family is the youngest within the Apiales order and originated in Australasia ~87 Ma. Apiaceae species are adapted to a wide range of conditions (from rocky to aquatic environments) and altitudes (from sea level to high mountains) but are predominant in temperate regions of Eurasia and North America and relatively rare in the humid tropics.

The family is organized into four main subfamilies, of which the Apioideae is the largest (2900 species and more than 400 genera). The remaining species are distributed in the Azorelloideae, Mackinlayoideae and Saniculoideae subfamilies or remain unclassified (approximately 200 species). Despite the presence of few large genera, each comprising from 130 to 250 species (such as *Eryngium*, *Bupleurum*, *Ferula* and *Seseli* genera), 85% is constituted by no more than ten species, and 25% by a single species (monospecific genera).

Many umbellifers are consumed as food, where the edible part is constituted by the root (e.g., *Daucus carota*), the leaf (e.g., *Apium graveolens)* or the seed (e.g., *Cuminum cyminum*). It is worth underlining how several species were widely consumed in the past, while they have now been abandoned, replaced by more common species, or confined to a niche cuisine. This is the case of *Smyrnium olusatrum* (commonly known as Alexanders), whose flavor is intermediate between celery and parsley and was greatly appreciated by Roman and medieval cuisines but has been progressively replaced by celery. For a comprehensive review of the nutraceutical properties of the main Apiaceae species, see. Some Apiaceae find application in the beverage industry, where they are used mainly as flavoring (e.g., *Pimpinella anisum* and *Angelica archangelica*, respectively employed for the preparation of Greek ouzo and Italian vermouth). Several species are used as laxatives, diuretics, stimulants or sedatives; meanwhile, several species are established remedies or traditional/folk therapies against conditions, including spasmodic gastro-intestinal pains, digestive disorders, inflammations of mucous membranes of the upper respiratory tract and chronic cough. Many experimental and biological investigations have been conducted to validate the antimicrobial and antifungal activities traditionally attributed to several Apiaceae species, such as *Carum carvi*, *Foeniculum vulgare*, *Coriandrum sativum* and *Pimpinella anisum*. Some genera are widely used as ornamental garden plants (i.e., *Heracleum* spp., *Angelica* spp., *Astrantia* spp.), whereas some others are currently exploited for industrial purposes in the cosmetics sector (*Coriandrum sativum*, *Anethum graveolens*) or for the production of gums and resins (*Ferula* spp., *Dorema ammoniacum*). Finally, it is worth noting that not all Apiaceae species are edible; among the most toxic taxa are *Conium* spp. (poison hemlock), *Cicuta* spp. (water hemlock) and *Aethusa cynapium* (poison parsley). Despite their toxicity, the insecticidal activity of *Cicuta virosa* and the analgesic and anti-inflammatory activity of *Conium maculatum* represent some possible applications.

The economic influence of a species is a pivotal force driving the development of new relevant knowledge. Indeed, the information available for economically relevant species is expected to be far more abundant than the data available for species of little interest. To this end, the present review aims to highlight both the lack and progress achieved in the genomics and transcriptomics of this family, with a particular emphasis on their applications for breeding purposes.

2. Plant Reproductive Strategies and Breeding Schemes in the Apiaceae Family

Apiaceae species reproduce sexually and are characterized by a common floral structure (simple or compound umbel); they also exhibit different crossing strategies because of the wide variation in the mode of pollination. Few cases of obligate outcrossing (i.e., strict allogamy) are known in dioecious (such as the Australian genera *Aciphylla* and *Anisotome)* and gynodioecious (i.e., female and hermaphroditic plants coexist, such as *Gingidia, Scandia* and *Lignocarpa)* taxa. An opposite situation is instead represented by the completely self-pollinated genus Scandix (to which the *S. pecten-veneris* leafy edible species belongs), where the anthers dehiscence occurs immediately above the stigma. Beyond these extreme cases, Apiaceae exhibits a combination of allogamous and autogamous behaviors with a pollination that is almost exclusively insect-dependent. In this case, plants are generally hermaphrodite (bisexual flowers), although several cases of andromonoecious plants (i.e., hermaphrodite flowers coexisting with male flowers on the same plant) are known, even in core species such as *Daucus carota, Coriandrum sativum* and *Pastinaca sativa*. In particular, it has been shown that the number of hermaphrodite flowers decreases progressively from the first-order umbel to the last, while male flower numbers follow an opposite path, decreasing from the last (where it could be up to 100%) to the first order umbel. Pollen produced by male flowers is mainly moved from pollinators to other plants, and only a small amount is used to pollinate hermaphrodite flowers of the same plants.

Self-incompatibility (SI) and male sterility (MS) represent the two most effective reproductive strategies to drastically reduce the occurrence of self-pollination. No cases of self-incompatibility have ever been reported in this family, except for a partial SI hypothesized for *Eryngium alpinum* and *Zizia* spp.

MS seems to be more common within the Apiaceae family. Cytoplasmic male sterility (CMS), a maternally inherited trait that is often associated with single-gene mutations in the mitochondrial genome, has been reported in *Daucus carota* but also in *Foeniculum vulgare*, *Apium graveolens* and *Pastinaca sativa*. In *Apium graveolens*, the discovery of CMS genotypes has been reported since the 1980s, and few F1 hybrids available on the market for celery and celeriac (*A. graveolens* var. *rapaceum)* were obtained using this reproductive barrier. However, information on breeding schemes is extremely vague. On the contrary, CMS mutants have been available and extensively exploited since the 1940s, and the most widely grown carrot cultivars are nowadays hybrids. Of the two types of CMS identified in carrot, the ‘petaloid sterile’ is preferred over the ‘brown anther’ for hybrid development. In this latter type of CMS, first stamens form and appear phenotypically normal while complete microspore abortion and brown anthers turning brown are later observed. In ‘petaloid sterile’ plants, the anthers are transformed into petaloid structures during their early development and do not produce pollen. In *Foeniculum vulgare* and *Pastinaca sativa*, the existence of CMS is documented, and several F1 hybrids are available on the market, but also, in this case, breeding data are quite scanty.

When a nuclear gene (called restorer gene) is able to restore fertility in CMS lines, this reproductive barrier assumes the name of cytoplasmic-genic male sterility (CGMS). So far, CGMS has only been detected and exploited for F1 hybrid production in carrots.

Finally, Nuclear or genic male sterility (NMS or GMS) - the result of a single recessive nuclear gene(s) (although a dominantly inherited pattern is also possible) - has been described in *Apium graveolens* and *Daucus carota*, but to date, the use of this type of male sterility for commercial purposes in Apiaceae is scarcely documented.

Based on the most recent literature review, for other important crops such as *P. crispum*, *C. carvi*, *C. cyminum*, and *C. sativum*, no source for MS has been found yet, making the establishment of F1 hybrid lines impossible.

Whatever the breeding scheme used, the selection of parental lines and the evaluation of the resulting offspring play a key role. This is pursued through the combined use of phenotypic and genotypic analyses.

3. Genomic Resources for Breeding Varieties

The constitution of new varieties benefits from the synergistic effect of marker-assisted breeding (MAB) techniques and conventional breeding programs. The use of molecular markers is a well-established practice to estimate the homozygosity and to genotype both the parents (to be crossed) and the resulting offspring. Markers are also widely used for the evaluation of DUS parameters (diversity, uniformity and stability) in the phase of variety constitution, for the legal registration of new varieties and as an effective tool for addressing legal disputes related to improper use of registered varieties. For this wide range of purposes, SSR represent a very attractive class of markers because of their reproducibility, co-dominant nature, locus-specificity and random genome-wide distribution.

The development and use of SSR markers within the Apiaceae family have made great strides and different methodologies have been used to develop this type of markers.

Species	Methodology	SSR Identified	SSR Validated	Samples Tested	PIC
Anethum graveolens	SSR transfer from *D. carota*	30 gSSR	15 gSSR	5	n.s.
Anethum sowa	gDNA-seq	48,951 gSSR	10 gSSR	n.s.	n.s.
Angelica biserrata	RNA-seq	8371 EST_SSR	17 EST-SSR	208	0.44–0.83
Angelica dahurica	RNA-seq	33,724 EST-SSR	10 EST-SSR	56	0.27–0.63
Angelica gigas	gDNA-seq	138,113 gSSR	36 gSSR	16	0.44–0.89
Apium graveolens	RNA-seq	80 EST-SSR	28 EST-SSR	31	0.06–0.72
Arracacia xanthorrhiza	biotinylated SSR primer	26 gSSR	14 gSSR	58	0.00–0.65
Bupleurum chinense	I-SSR	100 gSSR	19 gSSR	22	0.20–0.92
Bupleurum falcatum	gDNA-seq	91,377 EST-SSR	21 gSSR	n.s.	n.s.
Centella asiatica	data mining from EST-db	686 EST-SSR	18 EST-SSR	n.s.	n.s.
Coriandrum sativum	RNA-seq	9746 EST-SSR	76 EST-SSR	14	0.00–0.79
Cuminum cyminum	gDNA-seq	8086 gSSR	23 gSSR	30	0.03–0.70
Daucus carota	hybridization-based library	n.s.	196 gSSR	n.s.	n.s.
Foeniculum vulgare	gDNA-seq	103,306 gSSR	27 SSR	100	0.03–0.92
Heracleum spp.	ddRAD-seq	54 gSSR	19 gSSR	48	n.s.
Notopterygium incisum	RNA-seq	13,149 EST-SSR	19 EST-SSR	24	0.53–0.83
Notopterygium oviforme	gDNA-seq	793	17 gSSR	94	0.37–0.64
Oenanthe javanica	RNA-seq	1233 EST-SSR	n.s.	n.s.	n.s.
Pimpinella anisum	SSR transfer from *D. carota*	30 gSSR	16 gSSR	5	n.s.
Scaligeria lazica	gDNA-seq	1982	40 gSSR	40	0.37–0.84

The drop in sequencing costs has made the direct assembly of entire genomes and, thus, SSR development, more convenient in terms of time and cost. To date, there are only eight genome assemblies available for the Apiaceae. Of these, four genomes are finely assembled into pseudomolecules (i.e., chromosomes, *Apium graveolens, Daucus carota, Centella asiatica* and *Coriandrum sativum*); whereas the others (*Angelica gigas, Oenanthe javanica, Bupleurum falcatum and Foeniculum vulgare*) were incomplete and/or assembled at the scaffold level.

Common Name	Scientific Name	Chr Number	Genome Size (Mbp)		Assembly Level	Sequencing Platforms
			Estimated	Assembled		
Celery	*Apium graveolens*	2*n* = 22	n.s.	3332	Chr	Illumina Hiseq4000; PacBio Seq I; Hi-C; 10x Genomics
Carrot	*Daucus carota*	2*n* = 18	473	421	Chr	Illumina Hiseq2000, Sanger (BAC libraries), 454 GS FLX
Asiatic pennywort	*Centella asiatica*	2*n* = 18	430	430	Chr	10x Genomics, Hi-C, Illumina Hiseq X
Coriander	*Coriandrum sativum*	2*n* = 22	2130	2119	Chr	Illumina Miseq, PacBio Seq I, 10x Genomics, Hi-C
Korean angelica	*Angelica gigas*	2*n* = 22	2670	804	Scaff	Illumina Hiseq2500
Java waterdropwort	*Oenanthe javanica*	2*n* = 22	n.s.	1278	Scaff	Illumina Hiseq2500
Sickle hare's-ear	*Bupleurum falcatum*	2*n* = 16	2120	922	Scaff	Illumina Hiseq2000
Fennel	*Foeniculum vulgare*	2*n* = 22	1320	1010	Scaff	Illumina Hiseq2500

Overall, complete genome assemblies reduce the effort and time required for conventional MAB and MAS approaches. In fact, they represent a treasure trove of markers to be exploited for MAB purposes and a valuable opportunity to decipher hundreds of functional and regulatory networks. Genome assemblies, supported by RNA-seq experiments, make it feasible to identify and characterize the genes controlling the agronomic traits of interest and tagging molecular markers to be used for introgression practices.

4. Concluding Remarks

The Apiaceae family includes thousands of species; nonetheless, the vast majority of the genomic and transcriptomic data available concern a limited number of crop plant species of economic interest, mainly *D. carota* and *A. graveolens*. Few cases of male sterility are known, and they are exploited for the production of commercial F1 hybrids, but the information is quite scanty and obviously contained by breeding companies. For species with no known cases of male sterility, there are multiple alternatives to avoid or limit self-pollination, but most of them have been poorly taken into account. This would explain why phenotypic-aided recurrent selection and mass selection-based breeding programs are still widely used in several apiaceous crops.

In the meta-analysis of the molecular markers available for marker-assisted breeding, the situation is slightly better even if for some crops of agricultural interest (e.g., *Carum carvi*, *Pastinaca sativa* and *Pimpinella anisum*), panels of markers supporting varietal constitution are absent or insufficient. Although whole-genome sequencing is one of the best approaches for identifying markers, assembled genomes (fully or partially) are only available for eight species.

We believe that the serious shortcomings pointed out in this study will encourage future studies aimed at the development of next-generation molecular tools to be exploited in crop plants whose agronomic improvements are still based exclusively on conventional and suboptimal breeding programs.
