Chloroplast DNA: A Promising Source of Information for Plant Phylogeny and Traceability

Freitas AS¹, da Anunciação RR², D’Oliveira-Matielo CB² and Stefenon VM**

Abstract
Chloroplasts are organelles with a specific DNA (cpDNA), which are responsible by the energetic metabolism in eukaryotic cells from plants and algae. Their key role is to perform the photosynthesis, when the cell converts sunlight and carbon dioxide into glucose (energy) and oxygen. Studies aiming to understand how chloroplast works are important to highlight details about plant evolution. The new approaches in molecular biology have raised the knowledge about the chloroplast genome characteristics and opened a wide range of possible applications for the increased background. In this review, we discuss the use of chloroplast DNA for plant phylogeny and traceability. Considering the increasing need for food safety, identification of gene flow from genetically modified plants and for better tools to police investigation, many works have demonstrated the viability of using cpDNA for these purposes, either for phylogeny of land plants or traceability of food, genetically modified plants and illicit drugs.

Keywords
cpDNA; Food safety; Genetically modified plants; Traceability

Introduction
Chloroplasts are organelles responsible for the energetic metabolism in autotrophic eukaryotic cells that set up the gathering and conversion of light and carbon dioxide into glucose and oxygen. They are found in plants and algae, corresponding to a center of energy production for the cells [1]. Actually, these organelles are distinguished from other types of plastids by the presence of chlorophyll, being most concentrated in the leaf mesophyll [2].

The key role of chloroplasts is to conduct and perform the photosynthesis, process in which the pigment chlorophyll captures the energy from sunlight, and converts and stores it in ATP and NADPH molecules while freeing oxygen from water. Subsequently, ATP and NADPH are used to synthesize organic molecules from carbon dioxide [3]. Leaf guard cells of most plant species contain chloroplasts that possess an active electron transport chain, being capable of fixing carbon dioxide and play a fundamental role in stomatal opening and closing [2].

The origin of the chloroplasts is correlated to an endosymbiosis event in which a eukaryotic cell encompassed a photosynthetic bacterium and both eukaryotic cell and bacteria adapted themselves to live in symbiosis [1,4]. This theory supports and explains how chloroplasts have their own genome and transcription machinery, as well its specific genetic code, not linked to the nuclear genome.

On this context, studies that goal to elucidate mechanisms of regulation of chloroplast metabolism are very important to understand how this organelle performs its biological activities and guarantee the adaptive success of plants. In this direction, new approaches in molecular biology – such as the Next Generation Sequencing (NGS) – have increased quickly the knowledge about genome sequence and functions in a wide range of organisms from bacteria to humans and complex plants [5].

The first chloroplasts genome study was officially reported in 1986 with the publishing of the complete tobacco (Nicotiana tabacum) chloroplast genome [6]. The chloroplast DNA, often abbreviated as cpDNA, is composed by a single large circular DNA molecule; typically 120,000-170,000 base pairs long (Figure 1). They can have a contour length of around 30-60 micrometers, and have a mass of about 80-130 million daltons [7]. Chloroplasts sequences studies – as well as mitochondrial sequences – have taken prominence in works of plant evolution, phylogeny and traceability due to its small length, simple structure, maternal inheritance characters and conserved sequences [8].

The efficacy of DNA in relation to protein analyzes are already recognized as being applied to a variety of matrices and easily extracted with low biomass [9,10]. Currently, a number of existing biomarkers models (RAPDs, AFLPs, SSRs, SNPs and SCARs) are sources of potential DNA fingerprinting generators, creating a specific identification for each organism [11], and can identify the differences between organisms at different taxonomic levels, like genera or, more specifically, differentiating distinct varieties within the same species.

In this review, we present and discuss the main biotechnological resources for phylogeny and traceability of plants based on chloroplast DNA sequences analyses.

Phylogeny of plants
Noncoding sequences from the chloroplast genome are a primary source of data for molecular systematic, phylogeographic, and population genetic studies of plants. Even though still relatively little is known about levels of variation among different regions of the chloroplast genome, the facility of working with and the low cost for sequencing have increased the interest by this approach [7].

Due to previous related characteristics, genes presents in chloroplast genomes have shown a great potential for phylogenetic markers, such as performed by many works [7,12,13]. These efforts have contributed for a new comprehension of genetic plant relationship. Clearly, a phylogenomic approach based on complete chloroplast genomes has the potential to resolve some important taxonomic inconsistencies still discussed by taxonomists. As an example, we built a phylogenomic tree based on the complete sequences of chloroplast genomes from 19 species and clearly resolved the major taxonomic groups (Figure 2). Even the polytomy observed for the two species of Bryophyta may be an insight to review the origin of this group.
Figure 1: Gene map of the *Cannabis sativa* cv *Carmagnola* chloroplast genome. The red lines indicate the extent of the inverted repeats (IRa and IRb), which separate the genome into small (SSC) and large (LSC) single copy regions. The genes related to the photosystems A and B are shown as green blocks. The gene map was drawn using OG View online tools [9] using the complete sequence of cpDNA deposited in GenBank under number NC_026562.1.

Figure 2: cpDNA based phylogenomic analysis. Phylogenomic analysis of 19 plant species from different taxa based on the complete sequence of the chloroplast genomes deposited in the GenBank.
Traceability of food

Concern about food safety as well as its authenticity, quality and legitimacy has increased due to bioterrorism assessment and the emerging implications of outbreaks of foodborne diseases [13,14]. Techniques based on genetic markers and PCR assay promote the growing required safety in products that possesses in its composition any vegetal base [11,15,16].

Genetic traceability arises as a way to include information about the origin, materials or ingredients present in plant products, as well as to determinate the genetic makeup of the products [13]. These approaches allow authenticating and investigating the veracity of the kind of plant used in product design even as in various processing phases [17].

Currently, studies aiming this concern have been carried out [11,18]. Some researchers have already sought to fill gaps in the traceability of industrialized products and products that are not directly based on the use of plants. Making traceability a reality in the trophic chain, as well as highly processed products [15,19].

Traceability of genetically modified plants (GMPs)

Although intense research has been performed about the safety for the human health and/or for the environment – and no significant hazards connected to its use has been detected –, genetically modified plants (also know as transgenic plants) are a still polemic point of discussion [20]. Authors as Nicolia and colleagues show the current concern of the academic society regarding transgenic plants and their relation to the environment [21]. The need for security and the fear of topics such as gene flow, low biodiversity and indicative of tumors incidence [22] are points that generate doubts about the real security of implementation of this plants, even though nothing was proved and methods used in some studies have been criticized [23,24].

An interesting point of view is the use of genetic markers in the genome of GMPs to evaluate the dissemination of modified DNA. As many transformations of plants are performed in the chloroplast genome, such as for production of interest proteins [25,26], cpDNA markers are a very viable alternative to trace the dissemination of genetic material along a population of plants and prevent possible undesirable events.

The use of traceability is already performed in the European Union [27] and in the USA. However, due to the problems already mentioned, there is still no homogeneity regarding the marking of these products or the methods of analysis [20], demonstrating that many efforts are still necessary to a viable and confident method for GMP’s traceability. Although no standard method exists, the scenario is very optimistic.

Traceability of illicit drugs

Illicit drugs are a problem of public health in the world, mainly in developing countries [28]. In a context of scarcity of information about this theme, to know the geographical origin from where the drugs consumed by people came from becomes very important information for police and health agents. Trace the routes of trafficking may help competent authorities to elucidate and combat crimes related to drugs [29]. On the other hand, traceability of plants such as Cannabis sativa and Erythroxylum coca may also provide genetic information that differentiates plants cultivated for medicinal use, from plants cultivated for illicit use.

The prospection and application of cpDNA markers from plants as marijuana (Cannabis sativa L.) are a possible alternative to better map the origin of seized plants and drugs. No definitive study about this theme has been published yet, but some groups have tried to determinate the relationships between different varieties of Cannabis sativa. Gonçalves [30], for example, observed distinctions in genes from different Cannabis samples seized by the Brazilian Federal Police in this country, while a panel of nuclear SSR markers was tested for drugs seizures in Australia and USA [31-33].

The usefulness of chloroplast genome sequences for forensic proposes was recently demonstrated by our group. We found significant differences among Cannabis sativa cultivars based on plastidial DNA (Figure 3), showing that this approach based on cpDNA is a viable and reproducible tool for traceability of this species. Two cultivars designated for medicinal use due to low cannabidiol content (Carmagnola and Dagestani) grouped together, while cultivars Yoruba Nigeria and Cheungsam (which present high cannabidiol content and are cultivated and commercialized as drug) are in different clusters (Figure 3). This result suggests that even though the genes related to the cannabidiol are located in the nuclear genome, the selection of plants can also be recognized through the chloroplast genome.

Conclusions

Since the beginning of the novel sequencing techniques, much information about cpDNA, as well as its intraspecific relationships, offered a great contribution to the scientific world. In the last 30 years, such research allowed the development of technologies and tools that are rapidly advancing and improving global security and traceability in plant-based products as well as uncovering questions about the plant evolution and adaptation.

Based on these data, we show that using complete chloroplast genome arises as a promising tool for classification and differentiation of plants. More researches as well as the sequencing of more chloroplasts genomes are needed to a better understand about phylogenetic relationships among plants and for the establishment of standard protocols for DNA traceability. Nonetheless, existing studies already show an exciting scenario.
References