UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

DEPARTAMENTO DE BOTÂNICA

PROGRAMA DE PÓS-GRADUAÇÃO EM BOTÂNICA

Aplicação de *DNA Barcoding* para identificação de espécies pertencentes às tribos Sisyrinchieae e Tigridieae (Iridaceae)

Tese de Doutorado

Tiago Luiz da Silva Alves

Orientadora: Prof. Dra. Tatiana Teixeira de Souza Chies

Outubro de 2013

Agradecimentos

À minha orientadora Profa. Dra. Tatiana Chies por ter me acolhido prontamente em seu laboratório, por todos os ensinamentos, pela amizade e principalmente pela imensa paciência.

À Profa. Dra. Janete Fett por ter acreditado na minha capacidade quando eu mesmo havia desistido.

Ao Dr. Olivier Chauveau que auxiliou de forma decisiva em inúmeras etapas desta tese e contribuiu muito para minha formação.

A Prof. Dra. Lilian Eggers pelos valiosos ensinamentos em taxonomia de Iridáceas.

À minha colega Dra. Eudes Alves pela amizade e pelo treinamento em todos os procedimentos de rotina de nosso laboratório.

Às minhas grandes amigas Adriana e Marília pela amizade, incentivo e conselhos.

Às minhas amigas Ângela e Valerí, que me acompanham desde o tempo do mestrado.

Aos grandes amigos e companheiros de campo Eduardo Pasini e Leando Dal Ri.

Aos amigos Cassiano Welker e Jaqueline Durigon, que gentilmente contribuíram com amostras para este trabalho.

À Profa. Dra. Mara Ritter pela grande amizade, pelo apoio constante e por compartilharmos a mesma fascinação pelas plantas.

Aos Professores Dra. Silvia Miotto e Dr. Arthur Fett-Neto por terem sido professores excelentes de disciplinas muito úteis à minha formação.

Aos colegas de pós-graduação laboratório de biologia molecular Juliana Faccineto, Juliana Lustosa, Joana, Cristiane, Paula e Tamara.

Aos alunos de iniciação científica do laboratório de biologia molecular Juliana, Meire, Victória, Mariana, Raíssa e Daniela.

Ao casal de amigos do laboratório de micologia Larissa e Juliano.

Aos colegas taxonomistas Camila Carneiro, Fernanda, Thaíssa, Camila Inácio, Priscila, Édson e Fernando.

À Gabriela que sempre cuidou muito bem da limpeza de nossos laboratórios.

Aos meus amigos “não-botânicos” César Brum, Matheus Alves e Jonatan Pompermaier por me auxiliarem em algumas coletas.

Aos Prof. Nélson Matzembacher ao meu amigo Lino Brum Filho por terem gentilmente aberto as portas de suas propriedades para coleta de material.

À minha esposa Lauren, pelo amor, paciência, compreensão e auxílio em todas as etapas de minha pós-graduação.

Aos meus pais Aureci e Sandra que sempre me incentivaram e me ensinaram a importância do conhecimento.

Aos meus sogros Edite e Hermes pelo apoio constante.

Aos revisores anônimos do artigo referente ao capítulo 1.

À banca examinadora pela solicitude.

Ao CAPES pelo fornecimento da bolsa de pós-graduação.

Mad World

“All around me are familiar faces

Worn out places, worn out faces

Bright and early for their daily races

Going nowhere, going nowhere

Their tears are filling up their glasses

No expression, no expression

Hide my head i wanna drown my sorrow

No tomorrow, no tomorrow

And I find it kind of funny

I find it kind of sad

The dreams in which i'm dying

Are the best i've ever had

I find it hard to tell you

I find it hard to take

When people run in circles

It's a very, very

Mad world, mad world

Mad world, mad world ...”

Tears for Fears

Sumário

Resumo..1

Abstract..2

Introdução geral..5

DNA barcoding..5

A família Iridaceae..10

A tribo Sisyrinchieae e o gênero *Sisyrinchium*..16

A tribo Tigridieae..19

Objetivos..22

Referências..23

Capítulo 1: Species discrimination in Sisyrinchium (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus....................................34

Abstract..36

Introduction..37

Material and methods..40

Candidate barcodes and sampling..40

DNA isolation, PCR amplification and sequencing....................................41

Genetic distance-based method: the assessment of the barcoding gap.........41

DNA sequence similarity-based method: a distance comparison approach..41

Tree-based method: identification by clustering individual accessions........42

Results..43

Rates of PCR amplification and sequencing..43

Assessment of the barcoding gap.....44

Species discrimination.....45

Discussion.....46

Universality of the candidate barcodes.....46

A complex evolutionary context may hamper species identification.....47

The low identification ability of plastid barcodes.....49

The modest rate of resolution of the ITS region.....50

Combining different genomes and the utility of the Bayesian inference.....51

Conclusions.....52

Aknowledgements.....53

References.....54

Data accessibility.....62

Tables and figures.....63

Supplementary material.....71

Capítulo 2: Applying DNA barcoding to South-American Tigridieae (Iridaceae): restricted variation limits identification in related species.....80

Abstract.....82

Introduction.....83

Material and methods.....87

Taxon sampling.....87

DNA extraction, amplification and sequencing.....87

Data analyses.....88

Results.....89

Universality of PCR amplification and DNA sequencing............................89

Variability levels and barcoding gap..90

Species identification...92

Discussion...95

Universality levels among markers...95

Overall divergence and evaluation of the barcoding gap.............................96

Identification ability..98

Conclusions...99

References..100

Supplementary material...106

Conclusão...108

Considerações finais...109

Referências...113

Resumo

As técnicas de *DNA barcoding* (código de barras de DNA) têm como objetivo principal a identificação taxonômica de organismos através da amplificação e análise de sequências de DNA curtas, padronizadas e previamente definidas. Apesar do sucesso relativo desta abordagem em animais usando um único *locus*, a aplicação deste método em plantas apresenta menor capacidade de identificar espécies usando uma única região gênica, levando a necessidade de utilização de múltiplos *loci*. Além disso, ainda há certo debate sobre qual região gênica seria mais apropriada para o *DNA barcoding* em plantas, embora as regiões plastidiais *rbcL*, *matK* e o espaçador intergênico *trnH-psbA* juntamente com o espaçador intergênico nuclear do RNA ribossomal (ITS) sejam as mais comumente utilizadas até então. As tribos Sisyrinchieae e Tigridieae da família Iridaceae foram testadas de acordo com diferentes métodos e marcadores indicados para o *DNA barcoding* em plantas. Os resultados indicaram uma alta universalidade para membros da tribo Sisyrinchieae, mas também revelaram uma capacidade de identificação de espécies considerada baixa. Apesar disto, os espaçadores ITS foram indicados como a melhor sequência para *DNA barcoding* em Sisyrinchieae. Em Tigridieae, problemas inerentes ao sequenciamento impediram a utilização dos ITS em nossas análises. Assim, apenas marcadores plastidiais foram utilizados na tentativa de identificar espécies, apresentando novamente resultados modestos. A região gênica que atingiu maior capacidade de identificação em Tigridieae foi o gene *matK*. A incapacidade de se alcançar maiores taxas de identificação provavelmente está relacionada à complexa história evolutiva apresentada pelos grupos em análise. Este trabalho forneceu o primeiro conjunto significativo de dados de *DNA barcoding* aplicados a dois importantes grupos de Iridaceae de considerável

biodiversidade no Brasil. As tribos em análise apresentam espécies consideradas filogeneticamente próximas e são de difícil identificação devido a sua morfologia homogênea, principalmente em estado vegetativo, justificando plenamente o uso de métodos molecularespara a identificação taxonômica.

Palavras-chave: *rbcL*, *matK*, *trnH-psbA*, ITS, *Sisyrinchium*, *Herbertia*, *Cypella*, *Calydorea*.

Abstract

The main objective of DNA barcoding methods is the taxonomic identification of organisms by amplifying and analyzing short, standardized and previously defined DNA sequences. In spite of the relative success of this approach in animals using a single *locus*, the application of this method in plants has less ability to identify species using a single gene region, leading to the need of using multiple *loci*. Furthermore, there is still some debate concerning which gene region would be more suitable for DNA barcoding in plants, although the plastid regions *rbcL*, *matK* and the *trnH-psbA* intergenic spacer along with the nuclear intergenic spacer of ribossomal DNA (ITS) are the most commonly regions used thus far. The tribes Sisyrinchieae and Tigridieae of the family Iridaceae were tested according to different methods and markers used for DNA barcoding in plants. The results indicated a great universality for members of tribe Sisyrinchieae, but also showed a low ability to identify species. Nevertheless, ITS was imputed as the best sequence for DNA barcoding in Sisyrinchieae. In Tigridieae, problems inherent of ITS sequencing prevented its use in our analysis. Thus, only plastid markers were used in an attempt to identify species, showing modest results once again. The gene region that reached higher identification ability in Tigridieae was *matK*. The inability to achieve higher identification

levels is probably related to the complex evolutionary history presented by the groups in question. This work provided the first large data set of DNA barcoding applied to two important groups of Iridaceae with significant biodiversity in Brazil. The tribes in question present species considered phylogenetically related and are difficult to identify due to their homogeneous morphology, especially in vegetative stage, fully justifying the use of molecular methods for taxonomic identification.

Keywords: *rbcL*, *matK*, *trnH-psbA*, ITS, *Sisyrinchium*, *Herbertia*, *Cypella*, *Calydorea*.

Introdução Geral

Introdução geral

DNA Barcoding

A utilização de *DNA barcoding* (código-de-barras de DNA) como método para a identificação de organismos completou uma década de utilização em 2013. As premissas principais dessa metodologia consistem na amplificação via *Polymerase Chain Reaction* (PCR) de sequências de DNA definidas, curtas e padronizadas, como um novo paradigma na abordagem da identificação e do reconhecimento de espécies de qualquer organismo vivo (Hebert *et al*., 2003; Newmaster *et al*., 2006; Savolainen *et al*., 2005). A implementação e o desenvolvimento das técnicas de *DNA barcoding* têm mobilizado alguns grupos internacionais de cooperação, principalmente o *Consortium for the Barcode of Life* (CBOL) e o *International Barcode of Life* (iBOL), que procuram estimular a pesquisa científica nesta área. Essa abordagem vem se tornando cada vez mais popular (Kress & Erickson, 2007), principalmente propulsionada pelo relativo sucesso alcançado pelo *DNA barcoding* em organismos animais. Praticamente todos os estudos em *DNA barcoding* em animais utilizam como marcador exclusivo o gene mitocondrial que codifica para a subunidade *c* da citocromo oxidase (*CO1*) (Hebert *et al*., 2003; Hollingsworth *et al*., 2011). Infelizmente, um análogo em genomas vegetais tem se mostrado difícil de encontrar, sobretudo pela taxa de evolução lenta verificada para o genoma mitocondrial vegetal (Pennisi, 2007; Vijayan & Tsou, 2010; Yu *et al*., 2011). Assim, o sucesso nas taxas de amplificação e sequenciamento (universalidade), bem como a proporção de organismos

identificados (poder de discriminação) costuma ser substancialmente inferior em plantas quando comparado ao obtido para animais (Fazekas *et al*., 2009).

Para que uma determinada região genômica seja utilizada como *DNA barcoding*, esta deve ser eficiente quanto à universalidade dos iniciadores (*primers*) utilizados para a amplificação, qualidade e facilidade de amplificação, e também apresentar uma variabilidade ajustada a um alto poder discriminatório (Hollingsworth *et al*., 2011; Vijayan & Tsou, 2010). Para atingir taxas adequadas de identificação de espécies, é necessário que a variação nas sequências seja suficientemente alta entre espécies para separá-las, mas suficientemente baixa dentro de uma mesma espécie a fim de estabelecer um limite preciso de separação (*barcoding gap*) (Lahaye *et al*., 2008). Entretanto, é particularmente desafiador encontrar sequências relativamente conservadas para a síntese de *primers* universais que flanqueiem marcadores com suficiente variação para distinguir adequadamente as espécies analisadas (Zhang *et al*., 2013). Equacionar um balanço adequado entre universalidade e poder discriminatório é bastante difícil porque muitas vezes aumentar os níveis de discriminação implica diminuir a universalidade e vice-versa. Há uma preocupação evidente em uniformizar e padronizar protocolos de *DNA barcoding* bem como torná-los práticos e fáceis de reproduzir (Gonzalez *et al*., 2009). O genoma plastidial vegetal apresenta taxas evolutivas relativamente lentas (Lahaye *et al*., 2008), o que tem forçado uma reavaliação da utilização de uma única sequência para o *DNA barcoding* de plantas a despeito de infringir algumas das premissas apregoadas para estudos com *barcodes*. Contudo, há certo consenso de que um conjunto de algumas sequências é necessário para obter a variabilidade necessária para a aplicação de *DNA barcoding* (Hollingsworth *et al*., 2011; Lahaye *et al*., 2008). A utilização de um número muito extenso de *loci* pode provocar aumento de custos sem melhorar significativamente o desempenho

do método, já que a partir de uma determinada quantidade de *loci* utilizados os níveis de identificação costumam aumentar marginalmente (Hollingsworth *et al*., 2011; Lahaye *et al*., 2008). Isso demonstra que a utilização maciça de dados talvez não seja a melhor estratégia para superar as limitações até então verificadas para o *DNA barcoding* em plantas. No entanto, o surgimento de ferramentas de bioinformática mais potentes acopladas ao uso de novas técnicas de sequenciamento (*Next Generation Sequencing*, NGS) podem redirecionar os rumos do *DNA barcoding* (Bhargava & Sharma, 2013). Esse tipo de abordagem vem sendo utilizada com sucesso em estudos de "*DNA ultra-barcoding*", em que os objetivos incluem a identificação de linhagens e até mesmo indivíduos pertencentes a uma mesma espécie (Kane *et al*., 2012), necessitando de quantidades muito grandes de informação para alcançar a resolução necessária para separar organismos tão intimamente relacionados.

As sequências mais recomendadas pelo CBOL (CBOL Plant Working Group, 2009) para o *DNA barcoding* de espécies vegetais são regiões pertencentes aos genes *rbcL* e *matK* do genoma plastidial (Hollingsworth *et al*., 2011; Yu *et al*., 2011), ainda que se saiba das limitações apresentadas por tais sequências que costumeiramente apresentam baixo poder de identificação de espécies, sendo que o *locus matK* em específico apresenta limitações recorrentes para amplificação e baixa universalidade (Yu *et al*., 2011). Esses dois marcadores são denominados "*core barcodes*" de acordo com o CBOL e são indicados como *loci* de primeira escolha para *DNA barcoding* em vegetais. Uma série de marcadores moleculares alternativos tem sido usada como *loci* adicionais no intuito de suprir tais limitações, embora muitas vezes não aumentem substancialmente a capacidade de discriminação (Hollingsworth *et al*., 2011). O espaçador intergênico plastidial *trnH-psbA* é o marcador plastidial complementar (*supplementary barcode*) recomendado pelo CBOL. Entretanto, outras sequências localizadas no genoma plastidial como *rpoC1*, *rpoB*, *ycf5*,

atpF, *psbK-I*, *accD*, *atpF-H*, *ndhJ* e *psbI* (Fazekas *et al*., 2008; Hollingsworth *et al*., 2011; Vijayan & Tsou, 2010; Yu *et al*., 2011) também são alternativamente utilizadas, mas esses *loci* não são suficientemente informativos para serem utilizados sozinhos, levando à necessidade de combinar diferentes regiões para discriminar espécies (Yu *et al*., 2011). De modo alternativo, a utilização de sequências posicionadas no genoma nuclear, como os espaçadores internos transcritos do DNA que codifica RNA ribossomal (nrITS), têm se mostrado promissora, principalmente por apresentar variabilidade e poder discriminatório superior àqueles encontrados quando *loci* plastidiais são utilizados (Hollingsworth *et al*., 2011). Tem se verificado um uso crescente desses marcadores moleculares, sendo que os mesmos têm sido recomendados pelo CBOL como *barcode* suplementar juntamente com o *locus trnH-psbA* do genoma plastidial. Porém, sequências oriundas do genoma nuclear estão sujeitas a eventos indesejáveis como a presença de cópias parálogas que tipicamente podem comprometer a eficiência de amplificação ou ainda a interpretação dos resultados (Vijayan & Tsou, 2010). Além disto, a universalidade limitada apresentada em algumas circunstâncias e principalmente a amplificação cruzada com DNA contaminante de origem fúngica (Hollingsworth *et al*., 2011) têm sido determinantes para a não inclusão até o momento de sequências ITS como *core barcode* pelo CBOL. Diante desse contexto, encontrar marcadores universais para o *DNA barcoding* em plantas tem sido uma tarefa consideravelmente árdua para os pesquisadores.

Em um sentido amplo, a aplicação do *DNA barcoding* em espécies vegetais pode fornecer subsídios para a melhor compreensão das relações taxonômicas em níveis hierárquicos inferiores (sobretudo espécie), bem como auxiliar na identificação correta de espécies de forma prática e rápida (Hollingsworth *et al*., 2011). Assim, o *DNA barcoding* é bastante útil para avaliar o contexto taxonômico baseado em caracteres morfológicos. Os

resultados fornecidos por estudos de *DNA barcoding* podem contribuir na formulação (e reformulação) de hipóteses filogenéticas, de sistemas de classificação e também de limites taxonômicos (Seberg *et al*., 2003; Stace, 2005; Tautz *et al*., 2003). No entanto, os benefícios do uso de *DNA barcoding* vão muito além de sua mera utilização para a resolução de problemas de identificação de espécies no contexto taxonômico. Essa técnica se aplica a qualquer situação que se beneficie de um processo rápido e eficiente de identificação de espécies, incluindo a avaliação da biodiversidade (de Vere*et al*., 2012;Gonzalez *et al*., 2009; Kress *et al*., 2009; Lahaye *et al*., 2008; Parmentier *et al*., 2013; Tripathi *et al*. 2013), a aplicação em biologia forense (Eurlings *et al*., 2013; Meiklejohn *et al*., 2013; Yan *et al*., 2013) e a caracterização de plantas medicinais (Chen *et al*., 2010; Gao *et al*., 2010; Han *et al*.,2012; Kool *et al*. 2012; Li *et al*., 2011; Liu *et al*., 2012).

Desde o seu surgimento, um intenso debate tem se estabelecido sobre o valor científico e sobre a eficiência do *DNA barcoding* enquanto metodologia para identificação de organismos (Cameron *et al*., 2006; Dunn, 2003; Ebach & Holdrege, 2005; Hickerson *et al*., 2006; Lipscomb *et al*., 2003; Meyer & Paulay 2005; Packer *et al*., 2009; Pires & Marinoni, 2010; Wheeler, 2005). Boa parte da comunidade científica que tem rejeitado sistematicamente o método é composta por taxonomistas de diversas áreas que temem que métodos moleculares acabem ocupando o espaço antes exclusivo da taxonomia tradicional (Packer *et al*., 2009; Pires & Marinoni, 2010). Alguns argumentam que os custos do *DNA barcoding* são demasiadamente altos (Cameron *et al*., 2006) enquanto outros atacam a qualidade metodológica das análises (Meyer & Paulay, 2005). Além disto, há um sentimento de que o financiamento de projetos focados em taxonomia tradicional seja preterido em favor daqueles focados em *DNA barcoding* (Ebach & Holdrege, 2005). Gregory (2005) refuta esse último argumento ao mencionar que a maioria dos recusros

obtidos para a pesquisa com *DNA barcoding* é fornecida por instituições que originalmente não financiavam projetos com enfoque taxonômico tradicional. Embora os custos com *DNA barcoding* sejam ainda consideráveis, os novos métodos de sequenciamento de DNA (NGS) prometem reduzir consideravelmente tais custos (Kane *et al*., 2012), tornando o método progressivamente mais acessível. Além disso, os métodos usados em pesquisas de *DNA barcoding* têm se aperfeiçoado, alavancados pelo aprimoramento de ferramentas de bioinformática (Bhargava & Sharma, 2013) e do desenvolvimento de novas abordagens e metodologias para interpretação de dados moleculares (Bhargava & Sharma, 2013; Li *et al*., 2011; Zhang *et al*., 2011). A despeito dos debates acalorados acerca da real utilidade do método, a utilização de *DNA barcoding* tem sido cada vez mais frequente, já que agrega informações bastante úteis no contexto de uma taxonomia integrativa que busca combinar diferentes tipos de informação para reconhecer e circunscrever espécies (Dayrat, 2005; Padial *et al*., 2010; Pires & Marinoni, 2010).

A família Iridaceae

A família Iridaceae é uma família de monocotiledôneas pertencente a Liliales *sensu* Cronquist (1988). Devido às modificações proporcionadas pelo advento do uso de caracteres moleculares em filogenias, seu posicionamento foi alterado (APG II, 2003; APG III, 2009). Assim, atualmente Iridaceae passou a pertencer a Asparagales juntamente com outras grandes famílias botânicas como Orchidaceae e Asparagaceae (APG II, 2003; APG III, 2009). Iridaceae é considerada uma das mais biodiversas famílias entre as monocotiledôneas compreendendo cerca de 2030 espécies distribuídas dentro de 65 a 75 gêneros (Goldblatt *et al*., 2008). As principais sinapomorfias que caracterizam o grupo são

a presença de folhas equitantes e isobilaterais, cristais estiloides, ovário ínfero e flores com três estames (Reeves *et al*., 2001). Historicamente o grupo apresenta alguns problemas quanto à classificação em subfamílias e tribos, sendo o trabalho de Goldblatt (1990) a principal referência para o estabelecimento dos principais clados infrafamiliares existentes ainda hoje. Originalmente, Iridaceae subdividia-se em quatro subfamílias: Isophysioideae, Nivenioideae, Iridoideae e Ixioideae (Goldblatt, 1990), estabelecidas fundamentalmente em função de caracteres não-moleculares. Um melhor esclarecimento das relações taxonômicas da família passou a ser alcançado pela utilização de técnicas moleculares para inferir as relações filogenéticas. A utilização do *locus* plastidial *rps4* como base para a primeira filogenia molecular da família ajudou a esclarecer algumas relações até então intrigantes, como o estabelecimento da enigmática subfamília Isophysioideae como grupo irmão do restante da família (Souza-Chies *et al*., 1997). A utilização de um conjunto de quatro regiões plastidiais agregou subsequentemente mais resolução ao conhecimento evolutivo de Iridaceae (Reeves *et al*., 2001), culminando então no trabalho que propôs a subdivisão mais atual para a família (Goldblatt *et al*., 2008), que a divide em sete subfamílias: Isophysioideae, Patersonioideae, Geosiridoideae, Aristeoideae — estas três últimas criadas em função da comprovação da parafilia de Nivenioideae *sensu* Goldblatt (1990) — Nivenioideae, Crocoideae (antiga Ixioideae) e Iridoideae.

A subfamília Isophysioideae conta com apenas uma espécie: a australiana *Isophysis tasmanica* (Hook.) T. Moore, restrita à ilha da Tasmânia, como representante da uma linhagem única em Iridaceae devido à presença de ovário súpero, e não ínfero como nos demais membros da família (Goldblatt *et al*., 2008; Goldblatt & Manning, 2008; Reeves *et al*., 2001). Patersonioideae é uma subfamília que também conta com um único gênero (*Patersonia* R. Br.) distribuído predominantemente na Austrália, mas é substancialmente

mais diversificada que a subfamília anterior, já que possui aproximadamente 21 espécies (Goldblatt *et al*., 2008). De forma recorrente, Geosiridoideae também é uma subfamília que apresenta um único gênero, desta vez contando duas espécies (*Geosiris albiflora* Baill. e *Geosiris aphylla* Goldblatt & J.C. Manning) de distribuição restrita à ilha de Madagascar (Goldblatt & Manning, 2008). Do mesmo modo, Aristeoideae é uma subfamília que possui um único gênero (*Aristea* Aiton), mas difere das anteriores por possuir uma distribuição bastante ampla pelo continente africano (Madagascar e África sub-Sahariana) e maior biodiversidade, abarcando aproximadamente 55 espécies (Goldblatt & Manning, 2008; Reeves *et al*., 2001). A subfamília Nivenioideae até então abrangia os gêneros agora incluídos em Patersonioideae, Aristeoideae e Geosiridoideae, o que lhe rendia uma condição de subfamília parafilética. Excluídos tais táxons, ressurge uma Nivenioideae monofilética e exclusivamente distribuída na África (restrita à região florística do Cabo, África do Sul), mas de circunscrição menos abrangente que inclui somente os gêneros *Nivenia* Vent. (11 espécies), *Witsenia* Thunb. (uma espécie) e *Klattia* Baker (três espécies) (Goldblatt *et al*., 2008; Goldblatt & Manning, 2008).

As duas subfamílias restantes de Iridaceae são substancialmente mais biodiversas que as anteriores e merecem atenção proporcional a sua maior dimensão. De acordo com dados moleculares, a subfamília Crocoideae é monofilética (Goldblatt *et al*., 2008), representada por 29 gêneros e aproximadamente 1025 espécies de distribuição praticamente restrita ao continente Africano, sendo considerado um grupo morfologicamente coeso e bem sustentado por algumas sinapomorfias (Goldblatt & Manning, 2008). Crocoideae está dividida em cinco tribos (Goldblatt *et al*., 2008), ainda que alguns desses grupos careçam de melhor sustentação e estejam sujeitos a rearranjos no futuro. Além disso, algumas relações dentro dos gêneros ainda são pouco compreendidas devido à baixa coesão

morfológica observada entre seus representantes (Goldblatt *et al*., 2008). A espécie *Crocus sativus* L. é digna de nota especial, visto que seus longos estigmas avermelhados ricos em terpenoides voláteis possuem grande valor econômico devido ao seu uso como condimento, o conhecido açafrão (Srivastava *et al*., 2010). Podemos destacar como substancialmente diversificados em Crocoideae os gêneros *Watsonia* Mill. (51 espécies), *Lapeirousia* Pourret (42 espécies), *Gladiolus* L. (262 espécies), *Romulea* Maratti (92 espécies), *Crocus* L. (85 espécies), *Hesperantha* Ker Gawl. (82 espécies), *Babiana* Ker Gawl. ex Sims (90 espécies), *Ixia* L. (67 espécies) e *Dierama* K. Koch (44 espécies) (Goldblatt & Manning, 2008). A subfamília Crocoideae representa o ápice de um grande fenômeno de diversificação dentro do continente africano culminando na colonização de diferentes ambientes por essas plantas, resultando numa maior amplitude geográfica de distribuição desses táxons e influenciando seu processo de diversificação (Goldblatt *et al*., 2008).

Por sua vez, a subfamília Iridoideae (cinco tribos, 30 gêneros e cerca de 900 espécies) apresenta distribuição geográfica e origem diferentes de Crocoideae (Goldblatt, 1990; Goldblatt & Manning, 2008; Goldblatt *et al*., 2008). Há uma dicotomia evidente entre os táxons mais remotos da subfamília (tribo Irideae, por exemplo), cuja origem e distribuição predominante estão associadas ao "Velho Mundo", em contraposição aos táxons mais derivados, cujas evidências apontam para uma diversificação experimentada principalmente no continente americano (Goldblatt *et al*., 2008, Chauveau *et al*., 2011; Chauveau *et al*., 2012). Iridoideae é igualmente um grupo monofilético bem sustentado por um número representativo de sinapomorfias e que possui em sua maioria gêneros mais bem definidos e sustentados por caracteres morfológicos quando comparada a Crocoideae, ainda que nas tribos Tigridieae e Trimezieae verifique-se uma série de incongruências importantes a serem resolvidas (Goldblatt *et al*., 2008; Reeves *et al*., 2001; Lovo *et al*.,

2012). Os principais gêneros de Iridoideae em número de espécies são *Iris* L. (280 espécies, incluindo *Belamcanda* Adans.), *Moraea* Mill. *lato sensu* (198 espécies), *Sisyrinchium* L. (140 espécies) e *Tigridia* Juss. (55 espécies) (Goldblatt & Manning, 2008).

O conhecimento filogenético acerca de grupos infrafamiliares de Iridaceae ainda é incipiente e restrito a alguns táxons. Apenas quatro gêneros de Iridaceae já foram alvo específico de investigação filogenética por meio do uso de sequências de DNA: *Moraea* (Goldblatt *et al*., 2002; Loeuille *et al*., 2003), *Sysirinchium* (Chauveau *et al*., 2011), *Crocus* (Harpke*et al*., 2013; Petersen *et al*., 2008) e *Iris* (Cuenoud *et al*., 2002; Goldblatt & Mabberley, 2005; Kozyrenko *et al*., 2009; Makarevitch *et al*., 2003; Martinez *et al*., 2010; Wilson, 2004; Wilson, 2009; Wilson, 2011). A tribo Trimezieae e a subfamília Iridoideae no Novo Mundo também foram alvo de estudos de cunho filogenético revelando a presença de problemas de circunscrição em vários gêneros, tais como *Tigridia* (Chauveau *et al*., 2012; Rodriguez &Sytsma, 2006), *Cypella* Herb. (Chauveau *et al*., 2012), *Trimezia* Salisb. ex Herb., *Neomarica* Sprague e *Calydorea* Herb. (Lovo *et al*., 2012;). Uma fração substancial dos estudos de filogenia molecular em Iridaceae concentra-se no gênero *Iris*, que vem sendo sistematicamente estudado através de múltiplas abordagens com diferentes marcadores (Kozyrenko *et al*., 2009; Makarevitch *et al*., 2003; Martinez *et al*., 2010; Wilson, 2004; Wilson, 2009; Wilson, 2011) estabelecendo inclusive o posicionamento de *Belamcanda* dentro deste gênero (Goldblatt & Mabberley, 2005). O gênero africano *Moraea* (*peacock irises*) teve sua parafilia revelada pelo uso de *loci* plastidiais (Goldblatt *et al*., 2002; Loeuille *et al*., 2003). Já o gênero *Crocus* foi considerado monofilético de acordo com dois estudos de filogenia molecular, sendo que algumas redefinições em nível subgenérico foram sugeridas (Harpke *et al*., 2013; Petersen *et al*., 2008). Recentemente uma análise utilizando informações dos genomas nuclear, plastidial e mitocondrial

contemplou o gênero americano *Sysirinchium*, permitindo a reconstrução evolutiva do surgimento de tricomas produtores de óleo (elaióforos) como recompensa aos polinizadores e mostrando que esse processo possui influência marcante na evolução do gênero (Chauveau *et al*., 2011). De forma semelhante, o surgimento de elaióforos influenciou fortemente a diversificação de membros da subfamília Iridoideae no "Novo Mundo" (Chauveau *et al*., 2012). Uma análise combinando dados morfológicos e moleculares (plastidiais e nucleares) foi realizada para 32 espécies pertencentes à tribo Tigridieae (Iridoideae) apresentando novas propostas de circunscrição genérica e estabelecendo a monofilia do grupo (Rodriguez & Sytsma, 2006). Até o momento, a aplicação de *DNA barcoding* em Iridaceae foi realizada somente para o gênero *Crocus*, apresentando resultados modestos e confirmando a necessidade do uso de múltiplas regiões genômicas para aumentar a taxa de identificação de espécies (Seberg & Petersen, 2009).

O Brasil possui a flora mais biodiversa entre todos os países do globo (Giulietti *et al*., 2005) e abriga 165 espécies de Iridaceae distribuídas em 19 gêneros (Eggers *et al*., 2013). Há registro de uma quantidade significativa de espécies endêmicas no país, número que varia entre 70 (Eggers *et al*., 2013) e 68 espécies (Giulietti *et al*., 2005). Embora até o momento não se encontrem registros de espécies de Iridaceae em listas de espécies ameaçadas de extinção no Brasil, um número importante de espécies endêmicas apresenta alto risco de desaparecimento em função da ação antrópica sobre seu ambiente natural (Spier *et al*., 2008). O habitat preferencial das Iridaceae geralmente são os ambientes não florestais (Eggers, 2008; Goldblatt & Manning, 2008), representados no Brasil principalmente pelas áreas dos campos sulinos e do cerrado. Esses ambientes estão particularmente vulneráveis pela ação antrópica em decorrência da expansão da fronteira agrícola e urbanização experimentada atualmente no Brasil. O uso de *barcodes* para

auxiliar na identificação de espécies endêmicas ou daquelas alocadas em ecossistemas vulneráveis pode contribuir decisivamente para estratégias de conservação no país.

A tribo Sisyrinchieae e o gênero *Sisyrinchium*

A tribo Sisyrinchieae pertence à subfamília Iridoideae e conta com seis gêneros: *Libertia* Spreng., *Olsynium* Raf., *Orthrosanthus* Sweet, *Sisyrinchium*, *Solenomelus* Miers e *Tapeinia* Juss. (Goldblatt & Manning, 2008). O principal gênero desta tribo é *Sisyrinchium* que corresponde a mais de 75% da biodiversidade do grupo (Silvério *et al*., 2012). Algumas das principais características morfológicas vegetativas de *Sisyrinchium* incluem seu hábito herbáceo, com folhas de disposição equitante e com formas que vão de lanceoladas a lineares ou ainda ocasionalmente cilíndricas (Goldblatt *et al*., 1990; Goldblatt & Manning, 2008). A presença de um sistema de raízes fibrosas ou ainda a ocorrência de um rizoma caracterizam o sistema subterrâneo deste grupo de plantas (Goldblatt & Manning, 2008). A maioria das espécies é perene, mas algumas ocasionalmente são anuais sendo que *Sisyrinchium* é considerado o único gênero que apresenta este ciclo de vida em Iridaceae (Goldblatt & Manning, 2008). As inflorescências ocorrem em um escapo áfilo que abriga um agrupamento terminal de flores frequentemente envolvido por uma folha ou bráctea foliar (Goldblatt *et al*., 1990). Além disso, este escapo floral é tipicamente alado além de ser plano e comprimido (Karst & Wilson, 2012). As flores são actinomorfas e dialitépalas apresentando três estames, unidos ao menos parcialmente na base, além de ovário ínfero (Karst & Wilson, 2012). As flores apresentam valor ornamental com padrões de coloração extremamente variáveis que incluem combinações de tons de azul, violeta, rosa, branco e amarelo, ainda que tons de amarelo e azul predominem (Goldblatt *et al*., 1990; Goldblatt &

Manning, 2008). Há variação ocasional também em níveis intraespecíficos, como evidenciado na espécie *Sisyrinchium micranthum* Cav., que apresenta perigônios caracterizados por uma vasta gama de colorações e tamanhos (Tacuatiá *et al*., 2012a). Como citado anteriormente, a evolução e a diversificação do gênero *Sisyrinchium* possui forte influência da presença (ou ausência) de elaióforos florais secretores de óleos de origem lipídica que constituem adaptações associadas à ocorrência de diferentes tipos de sistemas de polinização (Chauveau *et al*., 2011). A ocorrência de elaióforos é recorrente em Iridaceae, onde *Sisyrinchium* assume posição de destaque por ser o maior gênero na família que apresenta este tipo de recompensa aos polinizadores, geralmente abelhas especialistas (Cocucci & Vogel, 2001).

Sisyrinchium (figura 1) é considerado um gênero bastante complexo (Chauveau *et al*., 2011), fato ilustrado pelas controvérsias quanto ao número de espécies do gênero, que varia consideravelmente conforme o autor e a época do estudo: cerca de 80 (Goldblatt *et al*., 1989), aproximadamente 140 (Goldblatt & Manning, 2008) ou ainda em torno de 200 (Rudall *et al*., 1986). A distribuição geográfica de *Sisyrinchium* é restrita ao continente americano (Karst & Wilson, 2012; Chauveau *et al*., 2011; Souza-Chies *et al*., 2012), e seus dois principais centros de diversidade encontram-se no México e na América do Sul (Goldblatt & Manning, 2008). As subdivisões infragenéricas baseadas em caracteres morfológicos propostas para o gênero também carecem de sustentação (Chauveau *et al*., 2011). Inicialmente o grupo foi dividido em dois subgêneros *Echtronema* e *Sisyrinchium* (Goldblatt *et al*., 1990). Alternativamente, uma série de trabalhos publicados por Ravenna (Ravenna, 2000, 2001, 2002, 2003a,b) subdividiu *Sisyrinchium* em oito seções como segue: *Echtronema*, *Hydastylus*, *Lenitium*, *Scirpeocharis*, *Segetia*, *Sisyrinchium*, *Spathirhachis* e *Viperella*. Nenhuma destas alternativas obteve plena confirmação nas filogenias existentes

para o gênero (Chauveau *et al*., 2011; Karst & Wilson, 2012), embora algumas das seções propostas por Ravenna tenham sido sustentadas (Chauveau *et al*., 2011).

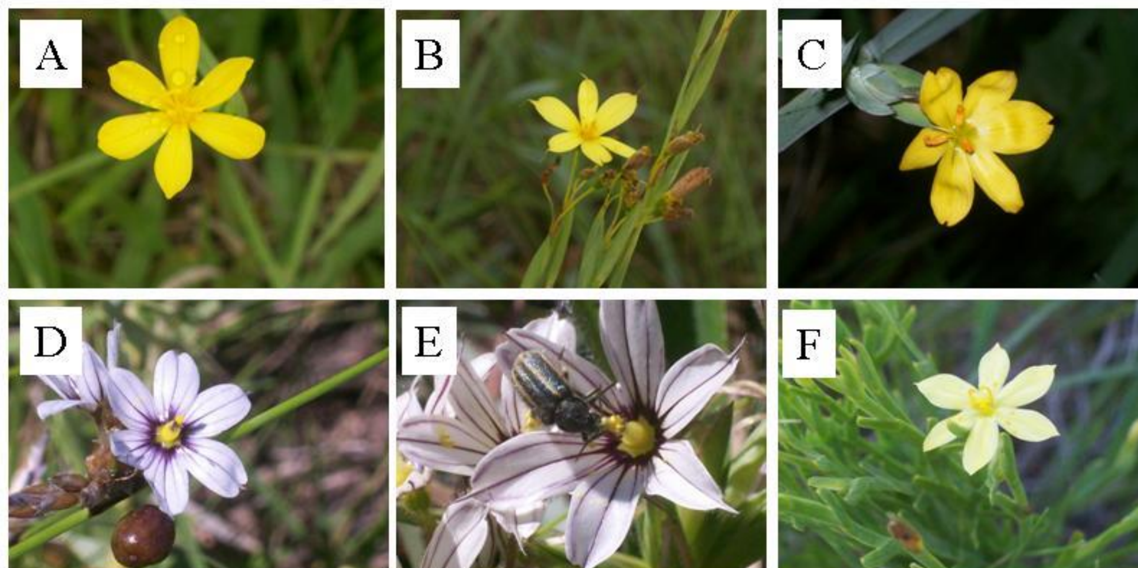


Figura 1: Representantes do gênero *Sisyrinchium*: A. *S. antemeridianum* (T.L.S.Alves135); B. *S. uliginosum* (T.L.S.Alves086); C. *S. palmifolium* (T.L.S.Alves225); D. *S. scariosum* (T.L.S.Alves079); E. *S. sellowianum* (T.L.S.Alves087) e F. *Sisyrinchium* sp. (T.L.S.Alves227).

A referida complexidade do gênero *Sisyrinchium* se deve fundamentalmente à dificuldade de circunscrever as espécies devido à escassez de caracteres diagnósticos e à

inerente variação de caracteres em populações da mesma espécie (Henderson, 1976). Várias espécies são absolutamente indistinguíveis quando em estado vegetativo. Além disso, a grande maioria das espécies apresenta florescimento fugaz, na ordem de dias, o que gera um obstáculo adicional à identificação taxonômica. A história evolutiva de *Sisyrinchium* é marcada por uma série de eventos que comumente confundem a identificação taxonômica baseada em morfologia e também as inferências filogenéticas moleculares. Hibridações naturais e poliploidia são fenômenos de ocorrência comum no grupo (Cholewa & Henderson, 1984; Henderson, 1976; Tacuatiá, 2012a,b; Yamaguchi & Hirai, 1987) que podem dificultar a elaboração de uma classificação taxonômica precisa. Novas espécies de *Sisyrinchium* foram descritas recentemente (Aita *et al*., 2013; Ceja-Romero *et al*., 2011) inclusive no Brasil (Aita *et al*., 2013), mostrando que ainda há muitas lacunas a preencher sobre a taxonomia do grupo. Iganci *et al*. (2011) demonstrou que cinco espécies de *Sisyrinchium* são endêmicas dos campos subtropicais de altitude do sul do Brasil, uma região muitas vezes negligenciada em termos de conservação e que apresenta considerável ação antrópica e expansão de atividades de silvicultura.

A tribo Tigridieae

A tribo Tigridieae (figura 2) pertence à subfamília Iridoideae e abrange cerca de 17 gêneros (Goldblatt & Manning, 2008) distribuídos em duas subtribos: Cipurineae e Tigridiineae (Goldblatt, 1990; Rodriguez & Sytsma, 2006). Entretanto, o número exato de espécies e gêneros desta tribo continua em aberto devido à existência de diversos grupos com problemas de circuncrição cuja definição ainda está em debate (Chauveau *et al*., 2012). As espécies desta tribo reunem uma série de características morfológicas comuns,

como a presença de bulbos subterrâneos e de uma estrutura floral especializada, onde os ramos do estilete estão intimamente associados aos estames (Rodriguez & Sytsma, 2006). No entanto, o que usalmente chama a atenção nesse grupo de plantas é a imensa diversidade de cores e formas de suas flores que inegavelmente atribuem potencial valor ornamental a tais espécies (Rodriguez & Sytsma, 2006). Sua distribuição se restringe às Américas, apresentando dois centros principais de biodiversidade: um na América do Sul e outro no México (Rodriguez & Sytsma, 2006).

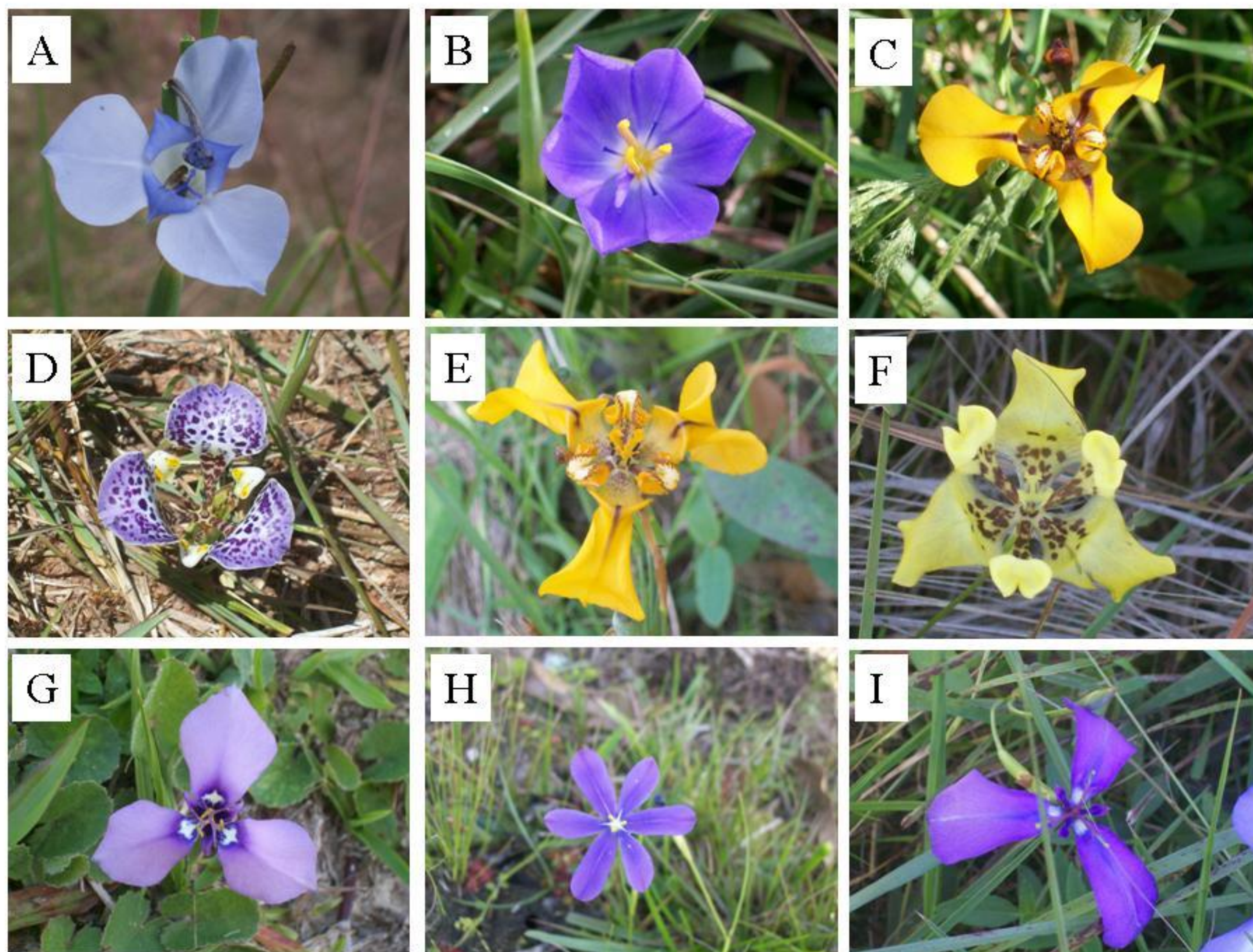


Figura 2: Representantes da tribo Tigridieae: A. *Gelasine coerulea* (TAL 134); B. *Calydorea crocoides* (T.L.S.Alves145); C. *Cypella herbertii* (T.L.S.Alves260); D. *Kelissa*

brasiliensis (T.L.S.Alves245); E. *Cypella fucata* (T.L.S.Alves244); F. *Cypella pusilla* (T.L.S.Alves197); G. *Herbertia* aff. *lahue* (T.L.S.Alves099); H. *Calydorea campestris* (T.L.S.Alves100) e I. *Herbertia pulchella* (T.L.S.Alves078).

Tigridieae foi analizada filogeneticamente por Rodriguez & Sytsma (2006) que, por meio do uso de marcadores nucleares, plastidiais e morfológicos, estabelecaram a monofilia do grupo e das duas subtribos propostas por Goldlatt (1990). No entanto, mais recentemente uma nova caracterização molecular do grupo usando exclusivamente dados plastidiais não confirmou a monofilia das subtribos Cipurineae e Tigridiineae, embora tenha demostrado mais uma vez a monofilia da tribo. De modo alternativo, Chauveau *et al*. (2012) classificaram Tigridieae em dois clados denominados “Clado A” e “Clado B”. Existem poucas discrepâncias em relação à classificação em subtribos de Goldblatt (1990), já que todo o “Clado A” é composto apenas de espécies alocadas em Cipurineae, mas alguns gêneros originalmente posicionados em Cipurineae (como *Gelasine* e *Eleutherine*) estão agora localizadas no “Clado B” juntamente com todos os gêneros de Tigridiineae (Chauveau *et al*., 2012).

Uma parcela significativa dos gêneros de Trigridieae possui importantes problemas de circunscrição (Chauveau *et al*., 2012). Alguns gêneros, como *Calydorea*, são obviamente polifiléticos (Chauveau *et al*., 2012), reunindo espécies cuja nomenclatura e circunscrição terão de ser obrigatoriamente revisadas. O gênero *Tigridia* representa um complexo (*Tigridia Complex*) de espécies cuja monofilia não foi sustentada, sendo que boa parte da variabilidade morfológica deste complexo de espécies pode ser atribuída a adaptações florais relacionadas a estratégias de polinização (Chauveau *et al*., 2012). A influência de estratégias de polinização na velocidade de diversificação já foi bem

caracterizadaa como um fator que pode propulsionar radiações adaptativas em membros de Crocoideae (Goldblatt & Manning, 2006) e em *Moraea* (Iridoideae) (Goldblatt *et al*., 2002).

A tribo Tigridieae pode ser considerada como significativamente complexa do ponto de vista nomenclatural devido aos já mencionados problemas de circunscrição genérica. Além disto, a diversificação de muitas linhagens em Tigridieae provavelmente está relacionadaà modificações nas estratégias de polinização (Chauveau *et al.,* 2012). Isto implica na hipótese de que provavelmente algumas linhagens de Tigridieae podem ter origem recente, e portanto caracterizar grupos de espécies estreitamente relacionadas, o que dificulta a resolução filogenética bem como a identificação por meio de *DNA barcoding*. Embora a presença de híbridos até o momento não seja descrita para Tigridieae, a ocorrência de citótipos hexaploides e octaploides já foi confirmada em *Herbertia lahue* (Stiehl-Alves, 2013), demonstrando que a evolução de algumas linhagens do grupo foram certamente influenciadas por processos de poliploidização.

Objetivos

Neste contexto em que: (i) o reconhecimento de espécies é bastante difícil, (ii) é ainda frequente a descrição de novas espécies e (iii) há ocorrência de espécies de distribuição restrita, a aplicação de técnicas de identificação moleculares como o *DNA barcoding* surge como uma ferramenta alternativa que pode auxiliar no reconhecimento e conservação destas espécies.

Este trabalho tem como objetivo geral utilizar os principais marcadores recomendados pelo CBOL para o *DNA barcoding* de plantas para identificar membros das

tribos Sisyrinchieae e Tigridieae. O desenvolvimento de novos métodos de identificação aplicáveis a estes grupos de plantas certamente contribuirá para o melhor entendimento da taxonomia destes organismos com consequências positivas para sua conservação, ecologia e utilização do ponto de vista econômico.

3. Referências

Aita AM, Chauveau O, Souza-Chies TT, Eggers L (2013) Two new species of *Sisyrinchium* (Iridaceae) from Subtropical Highland Grasslands of Southern Brazil. *Phytotaxa*, **88**, 9–18.

APG (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.

APG (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III.*Botanical Journal of the Linnean Society*, **161**, 105–121.

Bhargava M, Sharma A (2013) DNA barcoding in plants: Evolution and applications of in silico approaches and resources. *Molecular Phylogenetics and Evolution*, **67**, 631–641.

Cameron S, Rubinoff D, Will K (2006) Who will actually use DNA barcoding and what will it cost? *Systematic Biology*, **55**, 844–847.

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.

Ceja-Romero J, Espejo-Serna A, Lopez-Ferrari AR (2009) New species of *Sisyrinchium* (Iridaceae) from the state of Guanajuato, Mexico. *Acta Botánica Mexicana*, **87**, 83–90.

Chauveau O, Eggers L, Raquin C, Silvério A, Brown S, Couloux A, Cruaud C, Kaltchuk-Santos E, Yockteng R, Souza-Chies TT, Nadot S. (2011) Evolution of oil-producing trichomes in *Sisyrinchium* (Iridaceae): insights from the first comprehensive phylogenetic analysis of the genus. *Annals of Botany*, **107**, 1287–1312.

Chauveau O, Eggers L, Souza-Chies TT, Nadot S (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of Botany*, **110**, 713–729.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo, K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *Plos One***5**, e8613.

Cholewa AF, Henderson DM (1984) Biosystematics of *Sisyrinchium* section *Bermudiana* (Iridaceae) of the Rocky Mountains. *Brittonia*, **36**, 342–363.

Cocucci AA, Vogel S (2001). Oil-producing flowers of *Sisyrinchium* species (Iridaceae) and their pollinators in southern South America. *Flora*, **196**, 26–46.

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, **89,** 132–144.

Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407–415.

de Vere N, Rich TCG, Ford CR, Trinder SA, Long C, Moore CW, Satterthwaite D, Davies H, Allainguillaume J, Ronca S, Tatarinova T, Garbett H, Walker K, Wilkinson MJ (2012) DNA Barcoding the Native Flowering Plants and Conifers of Wales. *Plos One*, **7**, e37945.

Dunn, CP (2003) Keeping taxonomy based in morphology. *Trends in Ecology and Evolution*. **18**, 270–271.

Ebach MC, Holdrege, C (2005). More taxonomy, not DNA barcoding. *BioScience*, **55**,822–823.

Eggers L (2008) A família Iridaceae no Parque Estadual de Itapuã, Viamão, Rio Grande do Sul, Brasil. *Revista Brasileira de Biociências*, **6**, 167–175.

Eggers L, Chukr N, Lovo J, Gil A (2013) Iridaceae: in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro, Rio de Janeiro.

Eurlings MCM, Lens F, Pakusza C, Peelen T, Wieringa JJ, Gravendeel B (2013) Forensic Identification of Indian Snakeroot (*Rauvolfia serpentina* Benth. ex Kurz) Using DNA Barcoding. *Journal of Forensic Sciences*, **58**, 822–830.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy, DM, Hajibabaei, M, Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *Plos One***3**, e2802.

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S (2010) Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*, **130**, 116–121.

Giulietti AM, Harley RM, De Queiroz LP, Wanderley MDL, Van den Berg C (2005) Biodiversity and conservation of plants in Brazil. *Conservation Biology*, **19**, 632–639.

Goldblatt P, Mabberley DJ (2005) *Belamcanda* included in *Iris*, and the new combination *I. domestica* (Iridaceae: Irideae). *Novon*, **15**, 128–132.

Goldblatt P, Manning, JC (2008) The Iris family: natural history & classification. Timber Press.

Goldblatt P (1990) Phylogeny and classification of Iridaceae. *Annals of the Missouri Botanical Garden*, **77**, 607–627.

Goldblatt P, Rodriguez A, Powell MP, Davies TJ, Manning JC, van der Bank M, Savolainen V (2008) Iridaceae 'out of Australasia'? Phylogeny, biogeography, and divergence time based on plastid DNA sequences. *Systematic Botany* , **33**,495–508.

Goldblatt P, Savolainen V, Porteous O, Sostaric I, Powell M, Reeves G, Manning JC, Barraclough TG, Chase MW (2002) Radiation in the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution*, **25**, 341–360.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Petronelli P, Riera B, Roger A, Thebaud C, Chave J (2009) Identification of amazonian trees with DNA barcodes. *Plos One*, **4**, e7483.

Gregory, TR (2005) DNA barcoding does not compete with taxonomy. *Nature*, **434**, 1067.

Han, JP Shi, LC Chen, XC Lin, YL (2012) Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae. *Journal of Systematic and Evolution*, **50**, 227–234.

Harpke D, Meng SC, Rutten T, Kerndorff H, Blattner FR (2013) Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: Ancient hybridization and chromosome number evolution. *Molecular Phylogenetics and Evolution*, **66**, 617–627.

Hebert PDN, Cywinska A, Ball SL, DeWaard J (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Henderson DM (1976) A biosystematic study of Pacific Northwestern blue-eyed grasses (*Sisyrinchium*, Iridaceae). *Brittonia*, **28**, 149–176.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**, 729–739.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. *PlosOne*, **6**, e19254.

Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, Cronk Q (2012) Ultra-barcoding in cacao (*Theobroma* spp.) using chloroplast genomes and nuclear ribosomal. *American Journal of Botany*, **99**, 320–329.

Karst L, Wilson CA (2012) Phylogeny of the New World genus *Sisyrinchium* (Iridaceae) based on analyses of plastid and nuclear DNA sequence data. *Systematic Botany*, **37**, 87–95.

Kool A, de Boer HJ, Kruger A, Rydberg A, Abbad A, Bjork L, Martin G (2012) Molecular Identification of Commercialized Medicinal Plants in Southern Morocco. *Plos One*, **7**, e39459.

Kozyrenko MM, Artyukova EV, Zhuravlev YN (2009) Independent species status of *Iris vorobievii* NS Pavlova, *Iris mandshurica* Maxim., and *Iris humilis* Georgi (Iridaceae): evidence from the nuclear and chloroplast genomes. *Russian Journal of Genetics*, **45**, 1394–1402.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *Plos One*, **2**, e508.

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjurb O, Berminghamb E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot

in Panama. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18621–18626.

Iganci JRV, Heiden G, Miotto STS, Pennington T (2011) Campos de Cima da Serra: the Brazilian Subtropical Highland Grasslands show an unexpected level of plant endemism. *Botanical Journal of the Linnean Society*, **167**, 378–393.

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Li M, Cao H, But PPH, Shaw PC (2011) Identification of herbal medicinal materials using DNA barcodes *Journal of Systematic and Evolution*, **49**, 271–283.

Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 65–66.

Liu ZH, Zeng X, Yang D, Ren GM, Chu GY, Yuan ZR, Luo K, Xiao PG, Chen SL (2012) Identification of medicinal vines by ITS2 using complementary discrimination methods. *Journal of Ethnopharmacology*, **141**, 242–249.

Loeuille B, Goldblatt P, Manning JC, Savolainen V (2003) Molecular phylogeny of genus *Moraea* (Iridaceae: Irideae): DNA sequences analysis. *Acta Botanica Gallica*, **150**, 345–353.

Lovo J, Winkworth RC, Mello-Silva R (2012) New insights into Trimezieae (Iridaceae) phylogeny: what do molecular data tell us? *Annals of Botany*, **110**, 689–702.

Makarevitch I, Golvonina K, Scherbik S, Blinov A (2003) Phylogenetic relationships of the Siberian *Iris* species inferred from noncoding chloroplast DNA sequences. *International Journal of Plant Sciences*, **164**, 229–237.

Martinez J, Vargas P, Luceno M, Cuadrado A (2010). Evolution of *Iris* subgenus *Xiphium* based on chromosome numbers, FISH of nrDNA (5S, 45S) and *trnL-trnF* sequence analysis. *Plant Systematics and Evolution*, **289**, 223–235.

Meiklejohn KA, Wallman JF, Dowton M (2013) DNA Barcoding Identifies all Immature Life Stages of a Forensically Important Flesh Fly (Diptera: Sarcophagidae). *Journal of Forensic Sciences*, **58**, 184–187.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.

Monteiro SHN, Selbach-Schnadelbach A, de Oliveira RP, van den Berg C (2010) Molecular phylogenetics of *Galeandra* (Orchidaceae: Catasetinae) based on plastid and nuclear DNA sequences. *Systematic Botany*, **35**, 476–486.

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **84**, 335–341.

Packer L, Gibbs J, Sheffield C, Hanner R (2009) DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, **9**, 42–50.

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.

Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, Chuyong GB, Cruaud C, Hardy OJ (2013) How effective are DNA barcodes in the identification of African rainforest trees? *Plos One*, **8**, e54921.

Pennisi E (2007) Taxonomy - Wanted: A barcode for plants. *Science*, **318**,190–191.

Petersen G, Seberg O, Thorsoe S, Jorgensen T, Mathew B (2008) A phylogeny of the genus *Crocus* (Iridaceae) based on sequence data from five plastid regions. *Taxon*, **57**, 487–499.

Pires AC, Marinoni L (2010) DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. *Biota Neotropica*, **10**, 339–346.

Ravenna P. (2000) Revisional studies in the genus *Sisyrinchium* (Iridaceae) VI. *Onira Leaflets*, **5**, 13–19.

Ravenna, P (2001) The Iridaceae of the Cuyo region Argentina.*Onira Leaflets*, **6**, 1–18.

Ravenna P. (2002) Revisional studies in the genus *Sisyrinchium* (Iridaceae) IX. *Onira Leaflets*, **7**, 20–41.

Ravenna P. (2003a) Revisional studies in the genus *Sisyrinchium* (Iridaceae) X. *Onira Leaflets*, **8**, 20–24.

Ravenna P. (2003b) Revisional studies in the genus *Sisyrinchium* (Iridaceae) XI. *Onira Leaflets*, **8**, 48–54.

Reeves G, Chase MW, Goldblatt P, Rudall P, Fay MF, Cox AV, Lejeune B, Souza-Chies T (2001) Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany*, **88**, 2074–2087.

Rodriguez A, Sytsma KJ (2006) Phylogeny of the ‘"tiger-flower" group (Tigirideae: Iridaceae): molecular and morphological evidence. *Aliso*, **22**, 412–424.-

Rudall P, Kenton AY, Lawrence TJ (1986) An anatomical and chromosomal investigation of *Sisyrinchium* and allied general. *Botanical Gazette*, **147**, 466–477.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1805–1811.

Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a *Crocus*? *Plos One*, **4**, e4598.

Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology and Evolution*, **18**, 63–65.

Silvério A, Nadot S, Souza-Chies TT, Chauveau O (2012) Floral rewards in the tribe Sisyrinchieae (Iridaceae): oil as an alternative to pollen and nectar? *Sexual Plant Reproduction*, **25**, 267–279.

SouzaChies TT, Bittar G, Nadot S, Carter L, Besin E, Lejeune B (1997) Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene *rps4*. *Plant Systematics and Evolution*, **204**, 109–123.

Souza-Chies TT, Santos EK, Eggers L, Flores AM, Stiehl-Alves EM, Fachinetto J, Lustosa J, Corrêa LB, Tacuatiá LO, Piccoli P, Miz RB (2012) Studies on diversity and evolution of Iridaceae species in southern Brazil. *Genetics and Molecular Biology*, **35**, 1027–1035.

Stiehl-Alves EM (2013) Diversidade genética e evolução de espécies de *Herbertia* Sweet (Tigridieae: Iridaceae). Phd Thesis. Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 116pp.

Spier FF, Tacuatiá LO, Agostini G, Eggers L, Souza Chies TT (2008) Uso de marcadores PCR-RFLP como ferramenta na identificação de espécies da subfamília Iridoideae (Iridaceae) presentes no Parque Estadual de Itapuã, Viamão, RS, Brasil. *Revista Brasileira de Biociências*, **6**, 159–165.

Srivastava R, Ahmed H, Dixit RK, Dharamveer, Saraf AS (2010) *Crocus sativus* L.: A comprehensive review. *Pharmacognosy Review*, **4**, 200–208.

Stace CA (2005) Plant taxonomy and biosystematics - does DNA provide all the answers? *Taxon*, **54**, 999-1007.

Tacuatiá LO, Eggers L, Kaltchuk-Santos E, Souza-Chies TT (2012a) Population genetic structure of *Sisyrinchium micranthum* Cav. (Iridaceae) in Itapuã State Park, Southern Brazil. *Genetics and Molecular Biology*, **35**, 99–105.

Tacuatiá LO, Souza-Chies TT, Flores AM, Eggers L, Siljak-Yakovlev S, Kaltchuk-Santos E. (2012b) Cytogenetic and molecular characterization of morphologically variable *Sisyrinchium micranthum* (Iridaceae) in southern Brazil. *Botanical Journal of the Linnean Society*, **169**, 350–364.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.

Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, Roy S (2013) The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate *loci* for DNA barcoding of tropical tree species of India. *Plos One*, **8**, e57934.

Vijayan K, Tsou CH (2010) DNA barcoding in plants: taxonomy in a new perspective. *Current Science*, **99**, 1530–1541.

Wheeler QD (2005) Losing the plot: DNA 'barcodes' and taxonomy. *Cladistics*, **21**, 405–407.

Wilson CA (2004) Phylogeny of *Iris* based on chloroplast *matK* gene and *trnK* intron sequence data. *Molecular Phylogenetics and Evolution*, **33**, 402–412.

Wilson CA (2009) Phylogenetic relationships among the recognized series in *Iris* Section *Limniris*. *Systematic Botany*, **34**, 277–284.

Wilson CA (2011) Subgeneric classification in *Iris* re-examined using chloroplast sequence data. *Taxon*, **60**, 27–35.

Yamaguchi H, Hirai S (1987) Natural hybridization and flower color inheritance in *Sisyrinchium rosulatum* Bickn.. *WeedResearch*, **32**, 38–45.

Yan D, Luo JY, Han YM, Peng C, Dong XP, Chen SL, Sun LG, Xiao XH (2013) Forensic DNA Barcoding and Bio-Response Studies of Animal Horn Products Used in Traditional Medicine. *Plos One*, **8**, e55854.

Yu J, Xue JH, Zhou, SL (2011) New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution*, **49**, 76–181.

Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, Wan P, Feng J, Ward RD (2011) A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular Ecology*, **21**, 1848–1863

Zhang, W Fan, XH Zhu, SF Zhao, H Fu, LZ (2013) Species-Specific Identification from Incomplete Sampling: Applying DNA Barcodes to Monitoring Invasive *Solanum* Plants. *Plos One*, **8**, e55927.

Capítulo 1

Species discrimination in *Sisyrinchium* (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus

Artigo aceito para publicação na revista *Molecular Ecology Resources*

Species discrimination in *Sisyrinchium* (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus

TIAGO LUIZ DA SILVA ALVES, OLIVIER CHAUVEAU, LILIAN EGGERS and TATIANA TEIXEIRA DE SOUZA-CHIES*

Programa de Pós-Graduação em Botânica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, Av. Bento Gonçalves, 9500 - Prédio 43433, Brazil.

Keywords: ITS, *matK*, *trnH-psbA*, *Olsynium*, *Solenomelus*, *Orthrosanthus*

*Corresponding author: Tatiana Teixeira de Souza Chies, Fax: 51 3308 7755; e-mail: tatiana.chies@ufrgs.br

Abstract

DNA barcoding aims to develop an efficient tool for species identification based on short and standardized DNA sequences. In this study, the DNA barcode paradigm was tested among the genera of the tribe Sisyrinchieae (Iridoideae). *Sisyrinchium*, with more than 77% of the species richness in the tribe, is a taxonomically complex genus. A total of 185 samples belonging to 98 species of *Sisyrinchium*, *Olsynium*, *Orthrosanthus* and *Solenomelus* were tested using *matK*, *trnH-psbA* and ITS. Candidate DNA barcodes were analyzed either as single markers or in combination. Detection of a barcoding gap, similarity-based methods and tree-based analyses were used to assess the discrimination efficiency of DNA barcodes. The levels of species identification obtained from plastid barcodes were low and ranged from 17.35% to 20.41% for *matK* and 5.11% to 7.14% for *trnH-psbA*. The ITS provided better results with 30.61% to 38.78% of species identified. The analyses of the combined datasets did not result in a significant improvement of the discrimination rate. Among the tree-based methods, the best taxonomic resolution was obtained with Bayesian inference, particularly when the three datasets were combined. The study illustrates the difficulties for DNA barcoding to identify species in evolutionary complex lineages. Plastid markers are not recommended for barcoding *Sisyrinchium* due to the low discrimination power observed. ITS gave better results and may be used as a starting point for species identification.

Introduction

The identification of living organisms is often a challenging task to biologists and particularly to taxonomists. In an attempt to overcome this difficulty, DNA barcode has been largely used as a new biological tool to accurate and facilitate species identification (Herbert *et al*., 2003; Savolainen *et al*., 2005; Newmaster *et al*., 2006; CBOL Plant Working Group, 2009). DNA barcode uses short, variable and standardized DNA sequences easily amplified with universal primers (Vijayan & Tsou, 2010; Hollingsworth *et al*., 2011). The level of DNA barcode resolution in plants and animals is different (Pennisi, 2007; Fazekas *et al*., 2009). The mitochondrial gene Cytochrome Oxidase 1 (*cox*/CO1) in animal genomes is remarkably effective to discriminate species (Hollingsworth *et al*., 2011). An equivalent sequence has been hardly seen in plant genomes, since its mitochondrial sequences are too conservative to be used (Chase *et al*., 2005; Fazekas *et al*., 2008; Hollingsworth *et al*., 2011). Plant species might be naturally recalcitrant to DNA barcode methods due to incomplete lineage sorting, natural hybridization, polyploidy events and problems of taxon circumscription (Fazekas *et al*., 2009).

The most recommended sequences for DNA barcode in plants include the coding plastid regions *rbcL* and *matK*, usually named core barcodes (CBOL Plant Working Group, 2009), along with two supplementary non-coding regions: the plastid *trnH-psbA* intergenic spacer and the internal transcribed spacer (ITS) from the nuclear ribosomal DNA (Hollingsworth, 2011). There are evidences that *matK* and *trnH-psbA* are among the most variable markers in the plastid genome (cpDNA) and offer higher resolution than other cpDNA regions (Starr *et al*., 2009; Zhang *et al*., 2012). The nuclear ITS is currently considered to be the most promising sequence for barcode purposes because it usually

performs better when compared with either coding or non-coding plastid markers (Chen *et al*., 2010; Gao *et al*., 2010a,b; Pang *et al*., 2010; Liu *et al*., 2012). However, the ITS amplification is sometimes difficult to manage because of fungal contamination, presence of paralogous gene copies and polymorphic priming sites (Hollingsworth, 2011). DNA barcode approaches, which combine plastid and nuclear data, are generally most suitable to obtain adequate discriminating power (Hollingsworth, 2011; Hollingsworth *et al*., 2011).

The extent of the barcoding gap is a critical step in DNA barcode analysis because it estimates the magnitude between the means of intra- and interspecific distances (Meier *et al*., 2008). Overlap between the two distributions can be interpreted as a failure of DNA barcode (Meyer & Paulay, 2005). Different methods are used to ascribe species identification. Most of the DNA barcode approaches rely either on phylogenetic analyses such as distance matrix, maximum parsimony and probabilistic, or in algorithms based on comparison of genetic distances (Austerlitz *et al*. 2009; Zhang *et al*., 2011).

Sisyrinchieae is the largest tribe of the subfamily Iridoideae on the American continent (Goldblatt & Manning, 2008; Silvério *et al*., 2012). It comprises the genera *Libertia* Spreng., *Olsynium* Raf., *Orthrosanthus* Sweet, *Sisyrinchium* L., *Solenomelus* Miers and *Tapeinia* Juss. (Goldblatt & Manning, 2008). The number of accepted *Sisyrinchium* species varies from 80 (Goldblatt *et al*., 1989) to 200 (Rudall *et al*., 1986), but the most recent and reliable estimative considers approximatively 140 species (Goldblatt & Manning, 2008). This genus includes more than 77% of the species richness of the tribe Sisyrinchieae (Silvério *et al*., 2012). *Olsynium* is morphologically close to *Sisyrinchium*, both genera are monophyletic and form a strongly supported clade in a phylogenetic analysis using plastid, nuclear and mitochondrial data (Chauveau *et al*., 2011). *Sisyrinchium* is an herbaceous genus mainly characterized by the lanceolate to linear or

eventually terete leaves and the presence of either a rhizomatous rootstock or a fibrous root system (Goldblatt & Manning, 2008). A remarkably high morphological similarity is observed between closely related taxa (Souza-Chies *et al*., 2012), and such homogeneity leads to a scarcity of useful characters to ascribe reliable identification. Moreover, morphological traits can be fairly variable at the species level. The morphological plasticity observed for *Sisyrinchium micranthum* Cav., which exhibits three different morphotypes, is a striking example of this infraspecific variability (Tacuatiá *et al*., 2012a,b).

Recently, two main morphological infrageneric classifications were proposed. Goldblatt *et al*. (1990) divided the genus into two subgenera, *Echtronema* and *Sisyrinchium*, while eight sections named *Echtronema*, *Hydastylus*, *Lenitium*, *Scirpeocharis*, *Segetia*, *Sisyrinchium*, *Spathirhachis* and *Viperella* were retained as subdivisions in an alternative proposal (Ravenna, 2000, 2001, 2002, 2003a,b). The latest comprehensive phylogenetic analysis retrieved nine major clades, and most of the subdivisions previously recognized are not supported by the phylogeny (Chauveau *et al*., 2011).

The DNA barcode could constitute a suitable approach for species identification, even though it usually shows worse performance when tested in datasets rich in closely related taxa (Fazekas *et al*., 2009), and the low genetic divergence among barcode sequences may eventually culminate in a collapse of the barcoding gap (Hollingsworth 2011; Hollingsworth, *et al*., 2011). In an attempt to avoid this problem, three of the most polymorphic barcodes, ITS, *trnH-psbA* and *matK*, were selected (Hollingsworth, 2011).

The goal of this study was to test the DNA barcode paradigm for *Sisyrinchium* and its sister genus *Olsynium*. The main purposes of this study were to (i) estimate the magnitude of the barcoding gap for the three barcode regions selected; (ii) identify which

barcode region is more suitable to discriminate among species and individuals; (iii) compare the ability of distinct barcode methods to assign identification at the species level; and (iv) evaluate the current morphological classifications from a DNA barcode perspective.

Material and methods

Candidate barcodes and sampling

One nuclear (ITS) and two plastid (*matK* and *trnH-psbA*) *loci* were selected as candidate barcodes. The sampling included 185 accessions from 98 putative species including *Sisyrinchium* (n=88 spp.), *Olsynium* (n=7 spp.), *Solenomelus* (n=2 spp.) and *Orthrosanthus* (n=1 sp.). A broader sampling (n=16 individuals) was particularly intended for *S. micranthum* due to the polymorphism and remarkably wide distribution observed (Tacuatiá *et al*., 2012 a,b). Considering the number of accepted species proposed by Goldblatt & Manning (2008), the taxa sampling included 70% of the species richness of *Sisyrinchium*, covering the different subgeneric arrangements proposed in the literature. A broad intraspecific sampling is a critical step to increase the success of species identification and a special effort was made to fulfill such issue. Voucher information and GenBank accession numbers are given in Table S1 (Supporting Informations).

DNA isolation, PCR amplification and sequencing

DNA extraction procedures, PCR amplification and sequencing conditions were performed according to Chauveau *et al*. (2011). The primers used to amplify and sequence each barcode region are given in Table S2 (Supporting Informations).

Genetic distance-based method: the assessment of the barcoding gap

Alignment of the DNA sequences for single-barcode regions and its combinations were performed by muscle (Edgar, 2004), visually inspected and manually edited using MEGA5 (Tamura *et al*., 2011). The seven datasets resulting from each single region and every possible combination (Table 1) were evaluated using different DNA barcode methods. To estimate the barcoding gap, the relative distribution of pairwise genetic distances based on the Kimura-2-parameter (K2P) model of DNA substitution was performed using TAXONDNA software (Meier *et al*., 2006).

DNA sequence similarity-based method: a distance comparison approach

The TAXONDNA software (Meier *et al*., 2006) is based on genetic distances to perform taxonomic identification. It is based on an alignment-based parametric clustering algorithm able to compare each sequence to all others (Theodoridis *et al*., 2012). The K2P distance was chosen as model to the 'best match' (BM) and the 'best close match' (BCM) modes of the program. The 'best match' option searches a particular sequence through the closest match available in the barcode dataset, computing a correct assignment when the

taxa are the same. Two or more equally good matches classify the sample as ambiguous. A more stringent analysis is furnished by the 'best close match'. In this mode, a 95% pairwise distance threshold is calculated in order to establish a cut-off value when the distances exceed such parameter. In this case, the accessions are assigned to an extra category called 'no match' (Meier *et al*., 2006). All accessions of *Sisyrinchium*, *Olsynium*, *Orthrosanthus* and *Solenomelus* were tested, including those with single representatives. Evaluation of the discriminatory power of each region was measured with one-way ANOVA test followed by a *post-hoc* Tukey test for pairwise comparisons ($p \leq 0.05$) implemented in the PAST software (Hammer *et al*., 2001).

Tree-based method: identification by clustering individual accessions

The rate of monophyletic groups in phylogenetic trees is often appropriate to evaluate the discriminatory power of a DNA barcode *loci* (Maia *et al*., 2012; Theodoridis *et al*., 2012; Zhang *et al*., 2012). Therefore, four different phylogenetic methods were used. The neighbor-joining (NJ) trees were assessed applying MEGA5 (Tamura *et al*., 2011). NJ was performed running 1000 bootstrap replicates and using the K2P as the model of substitution for all barcode regions. A simple hierarchical clustering UPGMA based on the K2P model with 1000 bootstrap replicates was also conducted in MEGA5 since this method was used in some DNA barcode studies (Lahaye *et al*., 2008; Starr *et al*., 2009; Theodoridis *et al*., 2012). For the maximum parsimony (MP) analysis, two different softwares were used. A fast MP analysis was performed in MEGA5 with 100 random additions, with tree-bisection reconnection (TBR) and 1000 bootstrap replicates. Moreover, a Parsimony ratchet (PR) analysis (Nixon, 1999) was implemented with searches conducted according to the

method described in Chauveau *et al*., 2012. A Bayesian inference (BI) analysis was performed using MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) with parameter sets according to Chauveau *et al*., 2011. Only species with individuals forming a monophyletic group in the trees with a bootstrap value above 60% (Zhang *et al*., 2012) or posterior probability above 0.95 (Theodoridis *et al*., 2012) were accepted. The best trees generated in each analysis were used to draw hypothesis about species taxonomy using as reference the clades identified in Chauveau *et al*. (2011).

Results

Rates of PCR amplification and sequencing

All samples were successfully amplified and sequenced, except the ITS region for one accession of *Sisyrinchium rambonis* R.C. Foster (Table S1). The PCR amplification of the ITS region using the universal primers ITS5b and ITS4 did not reach universality. The visualization of the PCR products on a 1% agarose gel revealed two different amplification products for most of the samples. After purification and sequencing, the smaller fragment proved to be a contamination from fungal DNA. A specific forward primer (ITS38f) was used to amplify successfully the targeted region. The amplification and sequencing of plastid markers were fully accomplished (Table 1).

Assessment of the barcoding gap

The relative distribution of the frequencies from K2P distances (Fig. 1) showed two distinct patterns when plastid and nuclear barcodes were compared. The plastid regions *trnH-psbA* and *matK* presented a very low intraspecific variation since its distribution frequencies never exceeded a 0.5% distance. Moreover, the interspecific variation was also very low and overlapped substantially the intraspecific distance. The barcoding gap was not observed for these two barcodes probably due to the very low interspecific distances observed. The nuclear ITS distribution showed a distinct pattern. The intraspecific distance was high, reaching up to 2.0%, but the interspecific distance was superior as well and it did not fully overlap the intraspecific distribution.

Species discrimination

The comparison of the two outputs resulting from the BM and BCM modes revealed that BCM always recovered a lower identification rate (Table 2). The ITS single-region barcode outperformed the *matK* and *trnH-psbA* with an identification rate from four to five times higher. Although ITS was more effective, the 43.47% rate of correct identification in BCM mode was considered modest, distinguishing less than a half of the species. The identification rate obtained with both plastid regions was extremelly weak and consistent with the results obtained with the K2P distribution. The low level of divergence of the plastid barcodes in *Sisyrinchium* and related genera caused a substantial amount of ambiguous and incorrect identification in both BM and BCM modes (Table 2).

A non-linear pattern was observed when the barcodes were combined. The ITS + *trnH-psbA* barcodes showed a success of 48.36%, higher than the 43.47% obtained for ITS in BCM mode. On the other hand, in the ITS + *matK* combination this rate fell to 37.5%. *matK* + *trnH-psbA* combined presented additive effects when compared with plastid single-region separated and the resolution power was almost doubled in both BM and BCM modes. However, the combination of the three barcodes did not achieve better performance and the species identification rate (36.41%) remained slightly lower than the value observed with the combination of ITS + *matK* (37.50%) in BCM mode.

The tree-based methods performance was estimated in terms of species (Table 3) and individual identification (Table 4). Among the single-region analyses, the *trnH-psbA* performance was very poor, with assignments made properly only for 2.70-4.32% of the individuals, followed by the *matK* (15.14-17.30% of correct individual identification). The effectiveness of the nuclear single-region ITS was higher, reaching 27.17-34.24% of success. The combination of the two plastid *loci* provided a maximum success of individual identification of 19.46%. The two-region combinations ITS + *matK* and ITS + *trnH-psbA* performances were inferior or similar to the ITS performance in all tree-based analyses, except for BI, which reached 42.93% of individual identification. The same pattern was observed in the three-region combination, which outperforms all single and two-region barcode in BI trees, but in all other methods the single-region ITS was more effective than the ITS + *matK* + *trnH-psbA* combined dataset.

The DNA barcode performances observed for the single barcode analyses were similar regardless of the method used, but the BI outperformed the other methods when plastid and nuclear regions were combined. Furthermore, the BI trees obtained with the combination of nuclear ITS and plastid markers showed a better resolution of the deep

phylogenetic relationships with higher support at the internal nodes, particularly in the ITS + *matK* + *trnH-psbA* tree (Fig. 2). The ITS + *matK* + *trnH-psbA* dataset analyzed under BI achieved the best performance among the tree-based methods tested here, even though the overall identification rate remained modest.

The results obtained from the tree-based methods were congruent with those observed from TAXONDNA, but their identification ability was slightly lower (Table 4). The one-way ANOVA with *post-hoc* Tukey test ($p \leq 0.05$) showed that the mean value of identification was higher for datasets including the ITS barcode.

The analysis of the most reliable trees yielded information which can contribute to a better understanding of the taxonomy of *Sisyrinchium* and its allies. The BI tree using ITS + *matK* + *trnH-psbA* (Fig. 2) was the more efficient to assign species identification (Tables 3 and 4). The 12 clades depicted in that tree were consistent with the recent comprehensive phylogeny of the genus (Chauveau *et al*., 2011), with minor changes.

Discussion

Universality of the candidate barcodes

The *matK* and *trnH-psbA* markers reached the criteria of universal amplification and sequencing, however ITS sequences failed since a specific primer was required to prevent amplification of fungal DNA. Fungal contamination is considered to be one of the main potential problems for the use of ITS as DNA barcode (Hollingsworth *et al*., 2011) and it was clearly a challenge in the present study. The limitations of ITS with regard to the amplification and sequencing processes are well known (Hollingsworth *et al*., 2011) and

the amplification success rate attained for ITS is not uniform. Low amplification rates ranging from 41% (Gonzalez *et al*., 2009) to 75% (Gao *et al*., 2010a) were reported, but amplifications were also fully achieved in some studies (Roy *et al*., 2010; Yang *et al*., 2012). The use of specific primers is not desirable according to rigorous DNA barcode requirements, but it is necessary to reach higher amplification and sequencing rates in recalcitrant situations.

A complex evolutionary context may hamper species identification

The difficulty to assign species identification in DNA barcode studies can be related to the speciation patterns in the taxonomic group under study. The genus *Sisyrinchium* is highly complex due to several features, which include the lack of reliable morphological apomorphies (Henderson, 1976) and unresolved phylogenetic relationships at the species level (Chauveau *et al*., 2011; Karst & Wilson, 2012), but also reticulate evolution driven by polyploidy (Tacuatiá *et al*., 2012a,b) and hybridization (Henderson, 1976; Cholewa & Henderson, 1984; Yamaguchi & Hirai, 1987).

Furthermore, incomplete lineage sorting resulting from the retention of ancestral polymorphisms during the speciation processes is not uncommon among recently diverged species (Funk & Oamland, 2003; Blanco-Pastor *et al*., 2012). The most complete phylogenetic study of *Sisyrinchium* performed so far showed that diversification may have occurred rapidly in the clades I, IV, V, VI and IX (Chauveau *et al*., 2011). The putative occurrence of incomplete lineage sorting among these major clades may explain partly the low level of discrimination observed. Moreover, natural hybridization can reduce the efficiency of DNA barcode because it is prone to fail since it can hardly identify species in

the presence of extensive reticulation. Several North American taxa from Clade IX (e.g. *S. idahoense* E.P.Bicknell, *S. littorale* Greene and *S. montanum* Greene) present breeding systems permissive to hybridization (Henderson, 1976; Cholewa & Henderson, 1984) and reproductive isolation is not complete for species nested within Clade VII, such as *S. rosulatum* E.P.Bicknell and *S. iridifolium* Kunth (Yamaguchi & Hirai, 1987). These examples show that evolution through hybridization occurs among different *Sisyrinchium* lineages and probably contributes to the limited performance of DNA barcode. Lastly, polyploidy is also a speciation process that can drastically reduce the ability of DNA Barcode to discriminate plant species (Fazekas *et al*., 2009). Indeed, polyploids were detected in clades 0, I, II, VII and IX (Rudall *et al*.,1986), but the best studied example occurs in clade VII where three different ploidy levels were identified within *S. micranthum* (Tacuatiá *et al*., 2012a,b).

Taxonomic inconsistences are another issue that can contribute to confound DNA barcode results. In this sense, the systematics of *Sisyrinchium* remains poorly resolved, the total number of species is still controversial (Chauveau *et al*., 2011) and type specimens in herbaria are often unavailable (Karst & Wilson, 2012), leading to frequent misidentifications (Aita *et al*., 2013). Several North American species included in Clade IX are especially problematic with numerous nomenclatural conflicts (Ingram, 1968; Ward, 1968; Henderson, 1976; Cholewa & Henderson, 1984). Such problems emphasize the difficulty of species assignment based on morphological traits and may lead to a misinterpretation of the results obtained from molecular data.

The biogeographic patterns of dispersal and diversification of *Sisyrinchium* suggested that transitions towards elaiophores were associated with major biogeographical and ecological changes in the resulting clades (Chauveau *et al*., 2011). It implies that

closely related species are prone to share the same distribution area, hampering the accuracy of DNA barcode even in local approaches. As a result, there is no evidence that the DNA barcode performance in *Sisyrinchium* is substantially better within a particular clade or geographic area because the identification level is homogeneously low over the genus. The complex evolutionary history of *Sisyrinchium* does not seem to be restricted to particular clades or distribution regions, but encompasses probably the whole genus.

The low identification ability of plastid barcodes

The frequencies of K2P distances showed a wide overlap between intra- and interspecific distribution, particularly for plastid markers. The *matK* and *trnH-psbA* barcodes had no barcoding gap and a very low identification power. The lack of divergence and resolution presented by the *trnH-psbA* is unexpected for this non-coding spacer region. It might be related to the insertion of the coding region *rps19*, which is a specificity of numerous monocotyledons (Wang *et al*., 2008). Distribution of the ITS frequencies was charaterized by larger intraspecific distances resulting in a better resolution power for this *loci*. This distribution pattern was observed in other large genera such as *Ficus* L. (Moraceae) (Li *et al*., 2012a) and *Parnassia* L. (Parnassiaceae) (Yang *et al*., 2012). The smaller overlapping distributions of intra- and interspecific genetic distances do not characterize a perfect barcoding gap, but are associated to a better identification power.

It is generally admitted that DNA sequences in plants are usually more phylogenetically informative in the nucleus than in the plastid or mitochondrial genomes (Yao *et al*., 2010; Hollingsworth *et al*., 2011, Wang *et al*., 2011). The low level of DNA barcode success obtained with plastid markers is not uncommon in other plant genera,

especially in groups characterized by a complex taxonomy such as *Berberis* L. (Berberidaceae) (Roy *et al*., 2010), *Solanum* L. sect. *Petota* (Solanaceae) (Spooner, 2009), *Carex* L. (Cyperaceae) (Starr *et al*., 2009) and *Crocus* L. (Iridaceae) (Seberg & Petersen, 2009). Plastid barcodes such as *matK* and *trnH-psbA* are usually more efficient when applied to a sampling of distantly related species, common in floristic surveys for example (Lahaye *et al*., 2008; Burgess *et al*., 2011; Costion *et al*., 2011). The presence of numerous unresolved nodes and short branch lengths in the terminal clades of the phylogeny published by Chauveau *et al*. (2011) suggest that divergences have occurred recently in several lineages of *Sisyrinchium*. In this context, the plastid markers retained for DNA barcode studies might not be the best choice and there are little evidence that other plastid regions could bring better results as effective barcode for *Sisyrinchium* and related genera because the addition of extra information, especially from the same genomic origin, often brings only marginal increment to the resolution power (Fazekas *et al*., 2008; Seberg & Petersen, 2009). The *matK* and *trnH-psbA* sequences are not recommended, as single region or combined, to barcode *Sisyrinchium* and allies due to the absence of sufficient variation and low success of identification never exceeding 25.94% of the individuals and 23.47% of the analyzed species (Tables 3 and 4).

The modest rate of resolution of the ITS region

In spite of some amplification disadvantages, the nuclear ITS is regarded as one of the most appropriate DNA barcode regions because of its higher variability, which can distinguish even closely related species (Chen *et al*., 2010; Hollingsworth, 2011). The ITS K2P distance distribution showed less overlap between intra- and interspecific divergences

than plastid sequences (Fig. 1). ITS generally provides better identification rates than plastid data (Ren *et al*., 2010; Yang *et al*., 2012; Zhang *et al*., 2012), but such rates are usually higher than those obtained in this study for *Sisyrinchium* and related genera. Although ITS has achieved the best single-region barcode performance, less than a half of the individuals and species were identified. ITS was already successfully used to identify species of large genera. The identification rates using NJ method for *Parnassia* and *Lysimachia* L.(Myrsinaceae) reached respectively 84.4% (Yang *et al*., 2012) and 86.7% (Zhang *et al*., 2012). However, ITS can yield low resolution for taxonomic groups where hybridization and introgression occurred (Spooner, 2009; Roy *et al*., 2010; Schilling, 2011). The low level of identification observed within *Sisyrinchium* may be related to the complex evolutionary history of the genus. Thus, the use of single region ITS as DNA barcode in this group is not sufficient to get effective identification.

Combining different genomes and the utility of the Bayesian inference

The results obtained when the ITS was combined with the plastid markers were characterized by different rates of identification, depending on the method used for the analysis (Tables 2, 3 and 4). The combination of distinct markers generally increases the barcode performance, but such gain is often marginal and occasionally little barcode identification success is obtained (Hollingsworth *et al*., 2009). The combination of plastid markers with ITS in *Sisyrinchium* usually did not increase the identification power observed for the ITS alone, and even reduced it occasionally. This trend was observed for every combination including ITS whatever the tree-based method used, except for BI (Tables 3 and 4). Considering the similarity-based methods, an improvement of the

assignment success was observed only for the combination ITS + *trnH-psbA* (Tables 2 and 4). Some studies report that tree-based methods usually do not improve significantly the barcode performance (Lahaye *et al*., 2008; Roy *et al*., 2010). On the other hand, both MP and BI performed better when combined datasets were used to barcode members of Lamiaceae (Theodoridis *et al*., 2012). In the present study, the tree inferred with BI using the three combined DNA regions was considered the most reliable because it is not only the best resolved, but also because it recovered a phylogenetic tree quite consistent with Chauveau *et al*. (2011). Our results suggest that the NJ, UPGMA and MP analyses may be less appropriate to deal with ambiguity, especially in the context of recently diverged lineages. Since the combination of the three barcode candidates allows to identify almost 50% of the species and 43% of the individuals with BI, a conservative recommendation would be to use such method and markers as a starting point for barcoding *Sisyrinchium* and relatives, although alternative *loci* are necessary to achieve better discrimination.

Conclusions

Taxonomic identification at species level in *Sisyrinchium* might be fairly elusive due to a high genetic homogeneity of the DNA barcodes *loci* versus a fast diversification rate. The modest success of identification is certainly influenced by reticulation, incomplete lineage sorting, polyploidy and a challenging taxonomy. Considering these speciation events, DNA barcode candidates evaluated here did not distinguish species properly.

The ITS region can be used as a starting point to identify *Sisyrinchium* and relatives species because it is certainly the most promising sequence to accomplish DNA barcode objectives in closely related species up to now. A proposal suggesting the incorporation of

ITS into the CBOL core barcode for seed plants strengthens the growing importance of such marker (Li *et al*., 2011). Plastid barcodes yield little resolution and low variation rates up to now, although they may complement the information from nuclear ITS.

Looking for more informative DNA regions is a necessity to improve the DNA barcode efficiency. The multicopy nrDNA external transcribed spacer (ETS) region has been used as an alternative source of variation and may contribute to solve the lack of divergence (Logacheva *et al*., 2010; Schilling, 2011). The rise of the next-generation sequencing technologies and "ultra-barcoding" approaches (Kane *et al*., 2012) is likely to provide a massive amount of useful information. In this context, the development of new algorithms to improve the accuracy of species identifications would be one of the main goals for future DNA barcode studies (Li *et al*., 2012b; Zhang *et al*., 2011). However, the results of this work emphasize that speciation patterns in *Sisyrinchium* may represent a real challenge for DNA barcode application, since this approach has inherent limitations in the context of a complex evolutionary history.

Aknowledgements

This work received funding from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and was supported by the network "Bibliothèque du Vivant" funded by the CNRS, the Muséum National d'Histoire Naturelle, the INRA and the CEA (Centre National de Séquençage). The PhD scholarship of the first author was provided by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The authors thank Victória Muccillo for the english revision.

References

Aita AM, Chauveau O, Souza-Chies TT, Eggers L (2013) Two new species of *Sisyrinchium* (Iridaceae) from Subtropical Highland Grasslands of Southern Brazil. *Phytotaxa*, **88**, 9–18.

Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, Leblois R, Veuille M, Laredo C (2009) DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics*, **10**, S10.

Blanco-Pastor L, Vargas J, Pfeil BE (2012) Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean. *Linaria*. *Plos One*, 7, e39089.

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy, DM, Hajibabaei, Barrett, SCH (2011) Discriminating plant species in a local temperate flora using the *rbcL* plus *matK* DNA barcode. *Methods in Ecology and Evolution*, **2**, 333–340.

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society. Series B, Biological Sciences*, **360**, 1889–1895.

Chauveau O, Eggers L, Raquin C, Silvério A, Brown S, Couloux A, Cruaud C, Kaltchuk-Santos E, Yockteng R, Souza-Chies TT, Nadot S (2011) Evolution of oil producing trichomes in *Sisyrinchium* (Iridaceae): insights from the first comprehensive phylogenetic analysis of the genus. *Annals of Botany*, **107**, 1287–1312.

Chauveau O, Eggers L, Souza-Chies TT, Nadot S (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of Botany*, **110**, 713–729.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One*, **5**, e8613.

Cholewa AF, Henderson DM (1984) Biosystematics of *Sisyrinchium* section *Bermudiana* (Iridaceae) of the Rocky Mountains. *Brittonia*, **36**, 342–363.

Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A (2011) Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS One*, **6**, e26841.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One*, **3**, e2802.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.

Funk DJ, Omland, KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology*, *Evolution and Systematics*, **34**, 397–423.

Gao T, Yao H, Song J, Zhu Y, Liu C, Chen S (2010a) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evolutionary Biology*, **10**, 324.

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S (2010b) Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*, **130**, 116–121.

Goldblatt P, Henrich JE, Keating RC (1989) Seed morphology of *Sisyrinchium* (Iridaceae – Sisyrinchieae) and its allies. *Annals of the Missouri Botanical Garden*, **76**, 1109–1117.

Goldblatt P, Rudall P, Henrich JE (1990) The genera of the *Sisyrinchium* alliance (Iridaceae: Iridoideae): phylogeny and relationships. *Systematic Botany*, **15**, 497–510.

Goldblatt P, Manning JC (2008) The Iris Family: Natural History and Classification. Timber Press, Portland, Oregon.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C, Chave J (2009) Identification of Amazonian Trees with DNA Barcodes. *PLoS One*, **4**, e7483.

Hammer Ø; Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, **4**, 1–9.

Herbert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 313–321.

Henderson DM (1976) A biosystematic study of Pacific Northwestern blue-eyed grasses (*Sisyrinchium*, Iridaceae). *Brittonia*, **28**, 149–176.

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan RS, Chase MW, Gaudeul M, Hollingsworth PM (2009) Selecting barcoding *loci* for plants:

evaluation of seven candidate *loci* with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, **9**, 439–457.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One*, **6**, e19254.

Hollingsworth PM (2011) Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the USA*, **108**, 19451–19452.

Ingram, R (1968) Breeding barriers in some species of *Sisyrinchium*. *New Phytologist*, **67**, 197–204.

Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, Cronk Q (2012) Ultra-barcoding in cacao (*Theobroma* spp.) using chloroplast genomes and nuclear ribosomal. *American Journal of Botany*, **99**, 320–329.

Karst L, Wilson CA (2012) Phylogeny of the New World genus *Sisyrinchium* (Iridaceae) based on analyses of plastid and nuclear DNA sequence data. *Systematic Botany*, **37**, 87–95.

Lahaye R , van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA*, **105**, 2923–2928.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, **108**, 19641–19646.

Li HQ, Chen JY, Wang S, Xiong SZ (2012a) Evaluation of six candidate DNA barcoding *loci* in *Ficus* (Moraceae) of China. *Molecular Ecology Resources*, **12**, 783–790.

Li PC, Yu ZG, Han GS, Chu KH (2012b) Analyzing multi–locus plant barcoding datasets with a composition vector method based on adjustable weighted distance. *PLoS One*, **7**, e42154.

Liu Z, Zeng X, Yang D, Ren G, Chu G, Yuan Z, Luo K, Xiao P, Chen SJ (2012) Identification of medicinal vines by ITS2 using complementary discrimination methods. *Journal of Ethnopharmacology*, **141**, 242–249.

Logacheva MD, Valiejo-Roman CM, Degtjareva GV, Stratton JM, Downie SR, Samigullin TH, Pimenov MG (2010) A comparison of nrDNA ITS and ETS *loci* for phylogenetic inference in the Umbelliferae: An example from tribe Tordylieae. *Molecular Phylogenetics and Evolution*, **57**, 471–476.

Maia VH, Mata CS, Franco LO, Cardoso MA, Cardoso SR, Hemerly AS, Ferreira PC (2012) DNA barcoding Bromeliaceae: achievements and pitfalls. *PLoS One*, **7**, e29877.

Meier R, Kwong S, Vaidya G, Nq PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in the land plants: evaluation of *rbcL* in a multigene tiered approach. *Canadian Journal of Botany*, **84**, 335–341.

Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, **15**, 407–414.

Pang X, Song J, Zhu Y, Xie C, Chen S (2010) Using DNA barcoding to identify species within Euphorbiaceae. *Planta Medica*, **76**, 1784–1786.

Pennisi E (2007). Taxonomy. Wanted: a barcode for plants. *Science*, **318**, 190–191.

Ravenna P. (2000) Revisional studies in the genus *Sisyrinchium* (Iridaceae) VI. *Onira Leaflets*, **5**, 13–19.

Ravenna, P (2001) The Iridaceae of the Cuyo region Argentina. *Onira Leaflets*, **6**, 1–18.

Ravenna P. (2002) Revisional studies in the genus *Sisyrinchium* (Iridaceae) IX. *Onira Leaflets*, **7**, 20–41.

Ravenna P. (2003a) Revisional studies in the genus *Sisyrinchium* (Iridaceae) X. *Onira Leaflets*, **8**, 20–24.

Ravenna P. (2003b) Revisional studies in the genus *Sisyrinchium* (Iridaceae) XI. *Onira Leaflets*, **8**, 48–54.

Ren BQ, Xiang XG, Chen ZD (2010) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. *Molecular Ecology Resources*, **10**, 594–605.

Ronquist FR, Huelsenbeck JP (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Datt B, Bag SK, Singh PK, Nair NK, Husain T, Tuli R (2010) Universal plant DNA barcode *loci* may not work in complex groups: A case study with Indian *Berberis* species. *PLoS One*, **5**, e13674.

Rudall P, Kenton AY, Lawrence TJ (1986) An anatomical and chromosomal investigation of *Sisyrinchium* and allied general. *Botanical Gazette*, **147**, 466–477.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1805–1811.

Schilling, EE (2011) Hybrid genera in Liatrinae (Asteraceae: Eupatorieae). *Molecular Phylogenetics and Evolution*, **59**, 158–167.

Seberg O, Petersen G (2009) How many *loci* does it take to DNA barcode a *Crocus*? *PLoS One*, **4**, e4598.

Silvério A, Nadot S, Souza-Chies TT, Chauveau O (2012) Floral rewards in the tribe Sisyrinchieae (Iridaceae): oil as an alternative to pollen and nectar? *Sexual Plant Reproduction*, **25**, 267–279.

Souza-Chies TT, Santos EK, Eggers L, Flores AM, Stiehl-Alves EM, Fachinetto J, Lustosa J, Corrêa LB, Tacuatiá LO, Piccoli P, Miz RB (2012) Studies on diversity and evolution of Iridaceae species in southern Brazil. *Genetics and Molecular Biology*, **35**, 1027–1035.

Spooner DM (2009) DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *American Journal of Botany*, **96**, 1177–1189.

Starr JR, Naczi RF, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**, 151–163.

Tacuatiá LO, Eggers L, Kaltchuk-Santos E, Souza-Chies TT (2012a) Population genetic structure of *Sisyrinchium micranthum* Cav. (Iridaceae) in Itapuã State Park, Southern Brazil. *Genetics and Molecular Biology*, **35**, 99–105.

Tacuatiá LO, Souza-Chies TT, Flores AM, Eggers L, Siljak-Yakovlev S, Kaltchuk-Santos E. (2012b) Cytogenetic and molecular characterization of morphologically variable *Sisyrinchium micranthum* (Iridaceae) in southern Brazil. *Botanical Journal of the Linnean Society*, **169**, 350–364.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE (2012) DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources*, **12**, 620–633.

Vijayan K, Tsou CH (2010) DNA barcoding in plants: taxonomy in a new perspective. *Current Science*, **99**, 1530–1541.

Wang Q, Yu QS, Liu JQ (2011) Are nuclear *loci* ideal for barcoding plants? A case study of genetic delimitation of two sister species using multiple *loci* and multiple intraspecific individuals. *Journal of Systematics and Evolution*, **49**,182–188.

Wang RJ, Cheng CL, Chang CC, Wu CL, Tu TM, Chaw SM (2008) Dynamics and evolution of the inverted repeat-large single copy junction in the chloroplast genomes of monocots. *BMC Evolutionary Biology*, **8**, 36.

Ward, DB (1968) The nomenclature of *Sisyrinchium bermudiana* and related North American species. *Taxon*, **8**, 270–276.

Yamaguchi H, Hirai S (1987) Natural hybridization and flower color inheritance in *Sisyrinchium rosulatum* Bickn.. *Weed Research*, **32**, 38–45.

Yang JB, Wang YP, Möller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). *Molecular Ecology Resources*, **12**, 267–275.

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Up H, Zhu Y, Xiao P, Chen S (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One*, **5**, e13102.

Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, Wan P, Feng J, Ward RD (2011) A fuzzy-set-theory-based approach to analyze species membership in DNA barcoding. *Molecular Ecology*, **21**, 1848–1863.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Molecular Ecology Resources*, **12**, 98–108.

Data accessibility

The following data are given in the Supporting Informations available in the online version of this manuscript: taxonomic sampling and GenBank accession numbers (Table S1), primers used for amplification and sequencing (Table S2), number of samples per species (Table S3), nexus file containing the *matK* + *trnH-psbA* + ITS alignment (Appendix S1).

T.L.S.A and T.T.S.C. conceived and designed the study. O.C. and L.E collected the plant samples and performed the taxonomic identification. O.C. performed the molecular procedures. T.L.S.A. analyzed the data. T.L.S.A., O.C. and T.T.S.C. wrote the paper

Tables and figures

Table 1: Markers, sampling, alignment length, variability parameters and rates of amplification and sequencing for each DNA barcode marker.

Barcode region	*N* individuals	*N* species	*N* singletons	Alignment length (bp)	Variable characters, *N* (%)	PI characters, *N* (%)	Overall K2P distance, %	Success of amplification and sequencing, %
ITS	184	98	53	737	395 (53.60)	302 (40.98)	5.30	99.46
matK	185	98	52	1064	167 (15.70)	105 (9.87)	1.21	100
trnH-psbA	185	98	52	718	73 (10.17)	42 (5.85)	0.91	100
ITS + *matK*	184	98	53	1801	562 (31.20)	407 (22.60)	n/a	n/a
ITS + *trnH-psbA*	184	98	53	1455	468 (32.16)	344 (23.64)	n/a	n/a
matK + *trnH-psbA*	185	98	52	1782	240 (13.47)	147 (8.25)	n/a	n/a
ITS + *matK* + *trnH-psbA*	184	98	53	2519	635 (25.21)	449 (17.82)	n/a	n/a

n/a: not applicable. PI: parsimony informative.

Table 2: Number of sampled individuals, rates of sample identification and further parameters yielded by the 'best match' and 'best close match' functions of the TAXONDNA software for each DNA barcoding marker and combinations.

Barcode region	*N* individuals	Best match, *N* (%)			Best close match, *N* (%)				Threshold, %
		Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect	No match	
ITS	184	82 (44.56)	33 (17.93)	69 (37.50)	80 (43.47)	33 (17.93)	15 (8.15)	56 (30.43)	1.73
matK	185	26 (14.05)	135 (72.97)	24 (12.97)	22 (11.89)	105 (56.75)	45 (24.32)	13 (7.02)	0.09
trnH-psbA	185	18 (9.72)	153 (82.70)	14 (7.56)	16 (8.64)	135 (72.97)	4 (2.16)	30 (16.21)	0.15
ITS + *matK*	184	74 (40.21)	27 (14.67)	83 (45.10)	69 (37.50)	25 (13.58)	26 (14.13)	64 (34.78)	0.91
ITS + *trnH-psbA*	184	91 (49.45)	16 (8.69)	77 (41.84)	89 (48.36)	16 (8.69)	70 (38.04)	9 (4.89)	0.94
matK + *trnH-psbA*	185	48 (25.94)	101 (54.59)	36 (19.45)	43 (23.24)	89 (48.10)	31 (16.75)	22 (11.89)	0.10
ITS + *matK* + *trnH-psbA*	184	78 (42.39)	16 (8.69)	90 (48.91)	67 (36.41)	13 (7.06)	41 (22.28)	63 (34.23)	0.66

Table 3: Number of species and rates of species identification for each DNA barcoding marker and combinations using different tree-based methods.

Barcode region	*N* species	Correct identification, %				
		UPGMA	NJ	MP	PR	BI
ITS	98	30 (30.61)	31 (31.63)	32 (32.65)	38 (38.78)	33 (33.67)
matK	98	20 (20.41)	20 (20.41)	17 (17.35)	17 (17.35)	18 (18.37)
trnH-psbA	98	7 (7.14)	7 (7.14)	5 (5.11)	6 (6.12)	5 (5.11)
ITS + *matK*	98	28 (28.57)	27 (27.55)	26 (26.53)	37 (37.76)	47 (47.96)
ITS + *trnH-psbA*	98	31 (31.63)	29 (29.59)	33 (33.67)	30 (30.61)	44 (44.90)
matK + *trnH-psbA*	98	23 (23.47)	16 (16.33)	17 (23.47)	17 (17.35)	18 (18.37)
ITS + *matK* + *trnH-psbA*	98	33 (33.67)	38 (38.78)	29 (29.59)	33 (33.67)	50 (51.02)

Table 4: Comparison of sample identification rates, means and standards deviations (s.d.) for seven DNA barcoding candidates obtained by different methods. The letters indicate means significantly different according to the ANOVA with *post hoc* pairwise Tukey test.

Barcode region	*N* individuals	Correct identification, %							Mean ± s.d.
		UPGMA	NJ	MP	PR	BI	BM	BCM	
ITS	184	27.17	29.35	31.52	34.24	31.52	44.56	43.47	$34.55^{a} \pm 6.83$
matK	185	15.14	17.30	16.22	15.68	16.76	14.05	11.89	$15.29^{bc} \pm 1.84$
trnH-psbA	185	4.32	4.32	3.24	3.78	2.70	9.72	8.64	$5.25^{c} \pm 2.77$
ITS + *matK*	184	18.48	18.48	25.54	30.98	38.57	40.21	37.50	$29.97^{a} \pm 9.31$
ITS + *trnH-psbA*	184	27.17	24.46	32.07	29.89	38.57	49.45	48.36	$35.71^{a} \pm 10.03$
matK + *trnH-psbA*	185	19.46	12.97	16.22	15.68	15.14	25.94	23.24	$18.38^{b} \pm 4.72$
ITS + *matK* + *trnH-psbA*	184	23.91	26.63	24.46	26.63	42.93	42.39	36.41	$31.91^{a} \pm 8.43$

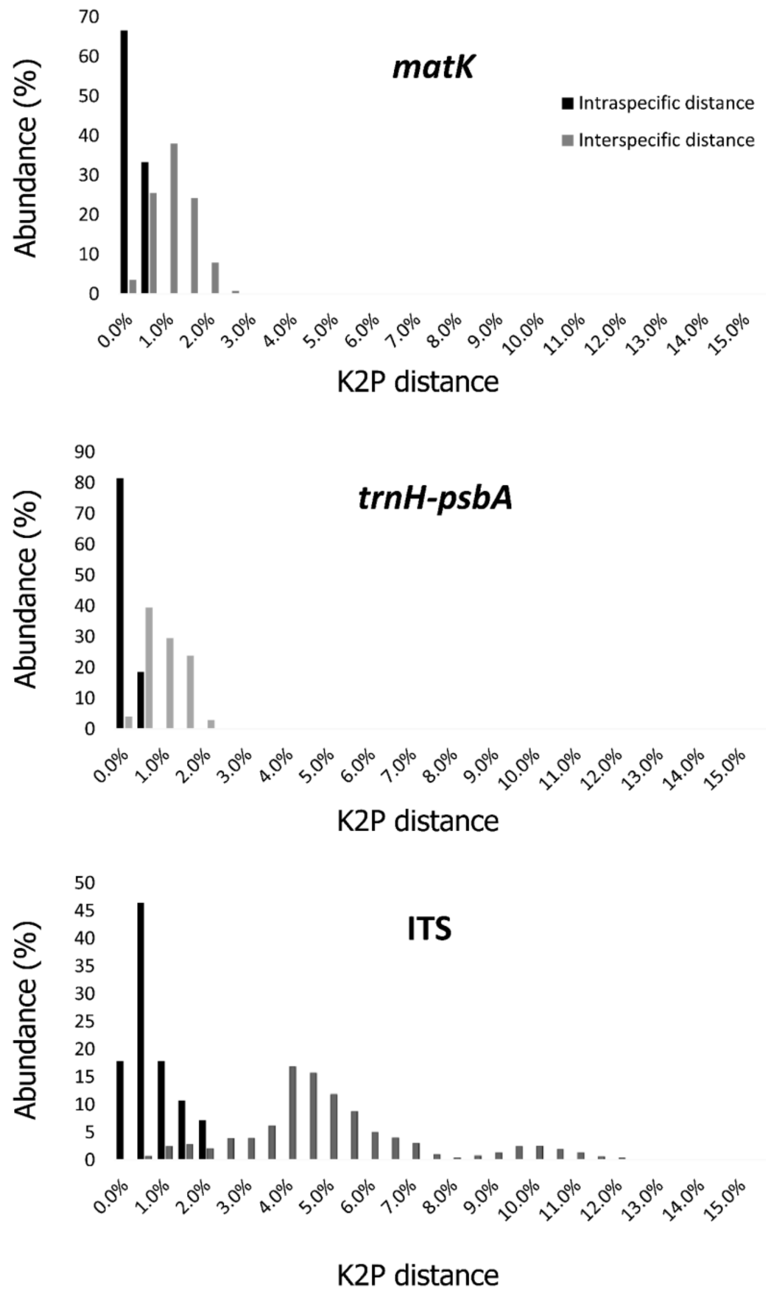
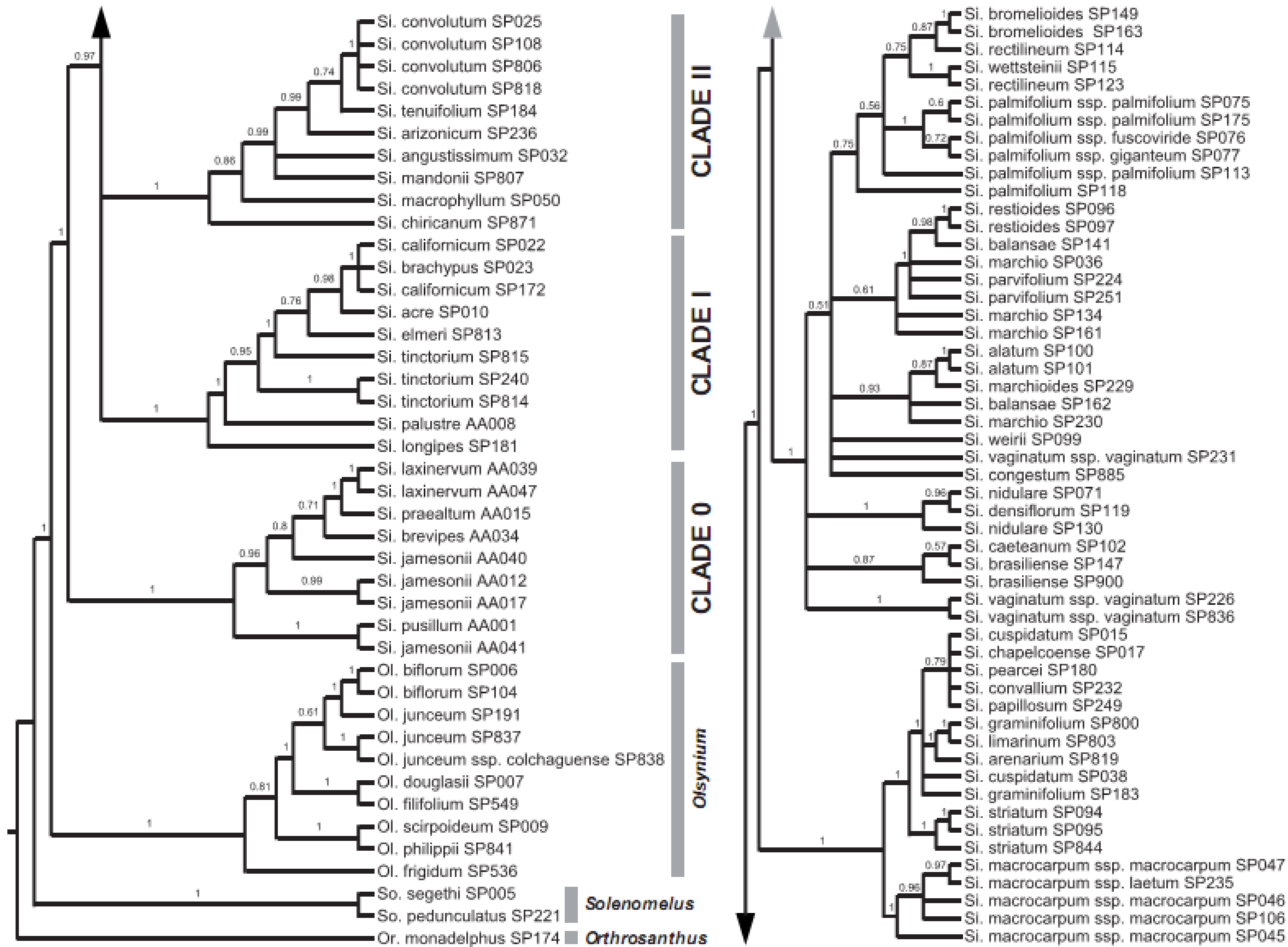


Figure 1: Histograms of the relative abundance of intra- and interspecific K2P distribution for the three DNA barcode candidate *loci*.



Si. convolutum SP025
Si. convolutum SP108
Si. convolutum SP806
Si. convolutum SP818
Si. tenuifolium SP184
Si. arizonicum SP236
Si. angustissimum SP032
Si. mandonii SP807
Si. macrophyllum SP050
Si. chiricanum SP871
Si. californicum SP022
Si. brachypus SP023
Si. californicum SP172
Si. acre SP010
Si. elmeri SP813
Si. tinctorium SP815
Si. tinctorium SP240
Si. tinctorium SP814
Si. palustre AA008
Si. longipes SP181
Si. laxinervum AA039
Si. laxinervum AA047
Si. praealtum AA015
Si. brevipes AA034
Si. jamesonii AA040
Si. jamesonii AA012
Si. jamesonii AA017
Si. pusillum AA001
Si. jamesonii AA041
Ol. biflorum SP006
Ol. biflorum SP104
Ol. junceum SP191
Ol. junceum SP837
Ol. junceum ssp. colchaguense SP838
Ol. douglasii SP007
Ol. filifolium SP549
Ol. scirpoideum SP009
Ol. philippii SP841
Ol. frigidum SP536
So. segethi SP005
So. pedunculatus SP221
Or. monadelphus SP174
CLADE II
CLADE I
CLADE 0
Olsynium
Solenomelus
Orthrosanthus
Si. bromelioides SP149
Si. bromelioides SP163
Si. rectilineum SP114
Si. wettsteinii SP115
Si. rectilineum SP123
Si. palmifolium ssp. palmifolium SP075
Si. palmifolium ssp. palmifolium SP175
Si. palmifolium ssp. fuscoviride SP076
Si. palmifolium ssp. giganteum SP077
Si. palmifolium ssp. palmifolium SP113
Si. palmifolium SP118
Si. restioides SP096
Si. restioides SP097
Si. balansae SP141
Si. marchio SP036
Si. parvifolium SP224
Si. parvifolium SP251
Si. marchio SP134
Si. marchio SP161
Si. alatum SP100
Si. alatum SP101
Si. marchioides SP229
Si. balansae SP162
Si. marchio SP230
Si. weirii SP099
Si. vaginatum ssp. vaginatum SP231
Si. congestum SP885
Si. nidulare SP071
Si. densiflorum SP119
Si. nidulare SP130
Si. caeteanum SP102
Si. brasiliense SP147
Si. brasiliense SP900
Si. vaginatum ssp. vaginatum SP226
Si. vaginatum ssp. vaginatum SP836
Si. cuspidatum SP015
Si. chapelcoense SP017
Si. pearcei SP180
Si. convallium SP232
Si. papillosum SP249
Si. graminifolium SP800
Si. limarinum SP803
Si. arenarium SP819
Si. cuspidatum SP038
Si. graminifolium SP183
Si. striatum SP094
Si. striatum SP095
Si. striatum SP844
Si. macrocarpum ssp. macrocarpum SP047
Si. macrocarpum ssp. laetum SP235
Si. macrocarpum ssp. macrocarpum SP046
Si. macrocarpum ssp. macrocarpum SP106
Si. macrocarpum ssp. macrocarpum SP045
CLADE IV
CLADE III

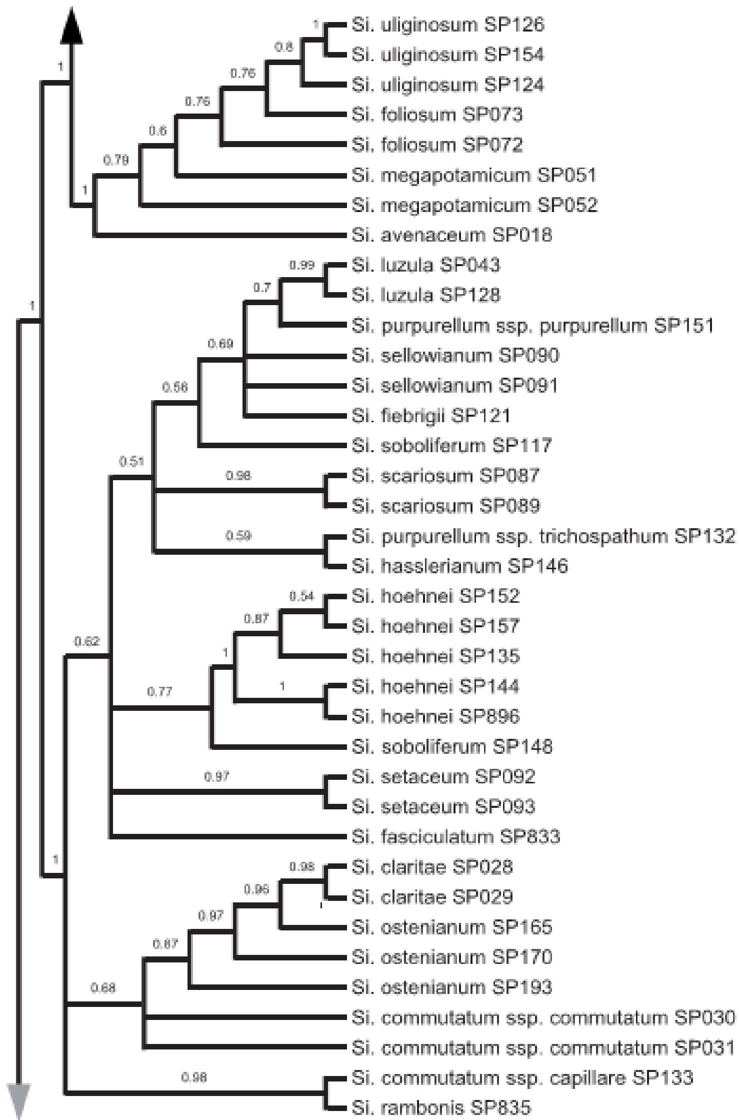
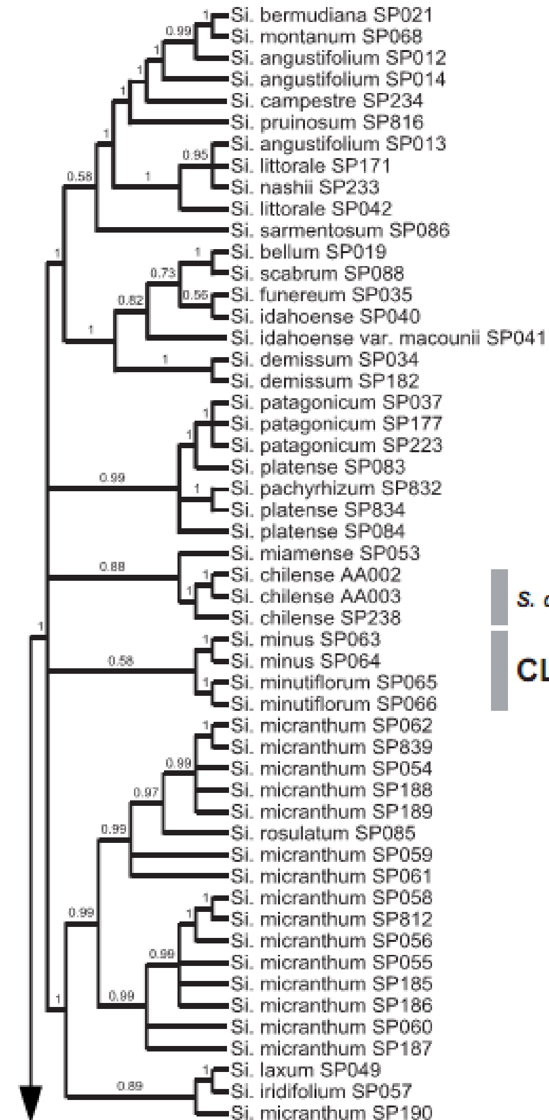

Si. uliginosum SP126
Si. uliginosum SP154
Si. uliginosum SP124
Si. foliosum SP073
Si. foliosum SP072
Si. megapotamicum SP051
Si. megapotamicum SP052
Si. avenaceum SP018
Si. luzula SP043
Si. luzula SP128
Si. purpurellum ssp. purpurellum SP151
Si. sellowianum SP090
Si. sellowianum SP091
Si. fiebrigii SP121
Si. soboliferum SP117
Si. scariosum SP087
Si. scariosum SP089
Si. purpurellum ssp. trichospathum SP132
Si. hasslerianum SP146
Si. hoehnei SP152
Si. hoehnei SP157
Si. hoehnei SP135
Si. hoehnei SP144
Si. hoehnei SP896
Si. soboliferum SP148
Si. setaceum SP092
Si. setaceum SP093
Si. fasciculatum SP833
Si. claritae SP028
Si. claritae SP029
Si. ostenianum SP165
Si. ostenianum SP170
Si. ostenianum SP193
Si. commutatum ssp. commutatum SP030
Si. commutatum ssp. commutatum SP031
Si. commutatum ssp. capillare SP133
Si. rambonis SP835
CLADE VI
CLADE V

Si. bermudiana SP021
Si. montanum SP068
Si. angustifolium SP012
Si. angustifolium SP014
Si. campestre SP234
Si. pruinosum SP816
Si. angustifolium SP013
Si. littorale SP171
Si. nashii SP233
Si. littorale SP042
Si. sarmentosum SP086
Si. bellum SP019
Si. scabrum SP088
Si. funereum SP035
Si. idahoense SP040
Si. idahoense var. macounii SP041
Si. demissum SP034
Si. demissum SP182
Si. patagonicum SP037
Si. patagonicum SP177
Si. patagonicum SP223
Si. platense SP083
Si. pachyrhizum SP832
Si. platense SP834
Si. platense SP084
Si. miamense SP053
Si. chilense AA002
Si. chilense AA003
Si. chilense SP238
Si. minus SP063
Si. minus SP064
Si. minutiflorum SP065
Si. minutiflorum SP066
Si. micranthum SP062
Si. micranthum SP839
Si. micranthum SP054
Si. micranthum SP188
Si. micranthum SP189
Si. rosulatum SP085
Si. micranthum SP059
Si. micranthum SP061
Si. micranthum SP058
Si. micranthum SP812
Si. micranthum SP056
Si. micranthum SP055
Si. micranthum SP185
Si. micranthum SP186
Si. micranthum SP060
Si. micranthum SP187
Si. laxum SP049
Si. iridifolium SP057
Si. micranthum SP190
CLADE IX
S. chilense group
CLADE VIII
CLADE VII

Figure 2: Bayesian inference tree for *Sisyrinchium* and related genera using the ITS + *matK* + *trnH-psbA* DNA barcode combination. Clade nomenclature follows Chauveau *et al*. (2011) with modifications. Abbreviations: Or, *Orthrosanthus*; So, *Solenomelus*; Ol, *Olsynium*; Si, *Sisyrinchium*.

Supplementary material

Table S1: Voucher information and GenBanK accession numbers of samples from *Sisyrinchium* and relatives.

Sample identification	Voucher	Accession numbers		
		matK	*trnH-psbA*	ITS
Olsynium biflorum SP006	*R. & E. Heaton OLS100.03* (living collection)	HQ606707	HQ606817	HQ607037
Olsynium biflorum SP104	*R. & E. Heaton OLS100.03* (living collection)	KF577237	KF577323	KF577162
Olsynium douglasii SP007	*R. & E. Heaton OLS100.02* (living collection)	HQ606708	HQ606818	KF577133
Olsynium filifolium SP549	*Chase I-243* (K)	KF577266	KF577353	KF577192
Olsynium frigidum SP536	*Chauveau H11006* (ICN)		KF577352	KF577191
Olsynium junceum SP191	*Chauveau & Aubert H09026* (ICN)	HQ606788	HQ606898	HQ607116
Olsynium junceum SP837	*Chauveau & Aubert H09050* (ICN)	HQ606813	HQ606923	HQ607141
Olsynium junceum ssp. *colchaguense* SP838	*Chauveau & Aubert H09051* (ICN)	HQ606814	HQ606924	HQ607142
Olsynium philippii SP841	*Chauveau H11022* (ICN)	KF577278	KF577365	KF577204
Olsynium scirpoideum SP009	*R. & E. Heaton OLS104.17* (living collection)	HQ606709	HQ606819	HQ607038
Orthrosanthus monadelphus SP174	*Chauveau & Pautz H09049* (ICN)	HQ606778	HQ606888	KF577180
Sisyrinchium acre SP010	*Chauveau & Heaton H09053* (ICN)	HQ606710	HQ606820	HQ607039
Sisyrinchium alatum SP100	*Eggers & Souza-Chies 232* (ICN)	HQ606756	HQ606866	HQ607085
Sisyrinchium alatum SP101	*Eggers & Souza-Chies* 239 (ICN)	KF577236	KF577322	KF577161
Sisyrinchium angustifolium SP012	*Chauveau H11012* (ICN)	KF577210	KF577295	KF577134
Sisyrinchium angustifolium SP013	*Chauveau H11013* (ICN)		KF577296	KF577135
Sisyrinchium angustifolium SP014	*Chauveau H09002* (ICN)	HQ606711	HQ606821	HQ607040
Sisyrinchium angustissimum SP032	*Chauveau H09008* (ICN)	HQ606722	HQ606832	HQ607051
Sisyrinchium arenarium SP819	*Chauveau H09047* (ICN)	HQ606809	HQ606919	HQ607137
Sisyrinchium arizonicum SP236	*Chauveau H09032* (ICN)	HQ606797	HQ606907	HQ607125
Sisyrinchium avenaceum SP018	*Eggers & Souza-Chies 280* (ICN)	HQ606714	HQ606824	HQ607043
Sisyrinchium balansae SP141	*Eggers & Souza-Chies 364* (ICN)	HQ606768	HQ606878	HQ607097
Sisyrinchium balansae SP162	*Eggers & Souza-Chies 408* (ICN)	KF577251	KF577337	KF577176
Sisyrinchium bellum SP019	*Chauveau & Heaton H09055* (ICN)	HQ606715	HQ606825	HQ607044
Sisyrinchium bermudiana SP021	*R. & E. Heaton SIS156.00* (living collection)	HQ606717	HQ606827	HQ607046
Sisyrinchium brachypus SP023	*R. & E. Heaton SIS102.01* (living collection)	HQ606719	HQ606829	HQ607048

Sisyrinchium brasiliensis SP147	*Eggers & Souza-Chies 379* (ICN)	HQ606770	HQ606880	HQ607099
Sisyrinchium brasiliensis SP900	*Eggers & Souza-Chies 642* (ICN)	KF577282	KF577370	KF577209
Sisyrinchium brevipes AA034	*Eggers et al. 034-2011* (ICN, CUZ)	KF153084	KF577290	KF577128
Sisyrinchium bromelioides SP149	*Eggers & Souza-Chies 382* (ICN)	KF577248	KF577334	KF577173
Sisyrinchium bromelioides SP163	*Eggers & Souza-Chies 410* (ICN)	HQ606776	HQ606886	HQ607105
Sisyrinchium caeteanum SP102	*Eggers & Souza-Chies 224* (ICN)	HQ606757	HQ606867	HQ607086
Sisyrinchium californicum SP022	*Chauveau H09006* (ICN)	HQ606718	HQ606828	HQ607047
Sisyrinchium californicum SP172	*UPSBG R08660* (living collection)	KF577255	KF577341	KF577179
Sisyrinchium campestre SP234	*Chauveau H09030* (ICN)	HQ606795	HQ606905	HQ607123
Sisyrinchium chapelcoense SP017	*Chauveau & Aubert H09004* (ICN)	HQ606713	HQ606823	HQ607042
Sisyrinchium chilense AA002	*Eggers et al. 002-2011* (ICN, CUZ)	KF153078	KF577284	KF577122
Sisyrinchium chilense AA003	*Eggers et al. 003-2011* (ICN, CUZ)	KF153079	KF577285	KF577123
Sisyrinchium chilense SP238	*Eggers, Chauveau & Heaton H09004* (ICN)	HQ606799	HQ606909	HQ607127
Sisyrinchium chiricanum SP871	*Chauveau H11024* (ICN)	KF577279	KF577367	KF577206
Sisyrinchium claritae SP028	*Eggers & Souza-Chies 267* (ICN)	HQ606720	HQ606830	HQ607049
Sisyrinchium claritae SP029	*Eggers & Souza-Chies 269* (ICN)	KF577213	KF577299	KF577138
Sisyrinchium commutatum ssp. *capillare* SP133	*Eggers & Souza-Chies 351* (ICN)	HQ606765	HQ606875	HQ607094
Sisyrinchium commutatum ssp. *commutatum* SP030	*Eggers & Souza-Chies 243* (ICN)	KF577214	KF577300	KF577139
Sisyrinchium commutatum ssp. *commutatum* SP031	*Eggers & Souza-Chies 245* (ICN)	HQ606721	HQ606831	HQ607050
Sisyrinchium congestum SP885	*Eggers & Souza-Chies 597* (ICN)	KF577280	KF577368	KF577207
Sisyrinchium convallium SP232	*Chauveau & Aubert H09028* (ICN)	HQ606793	HQ606903	HQ607121
Sisyrinchium convolutum SP025	*Chauveau s.n.* (ICN 190698)	KF577212	KF577298	KF577137
Sisyrinchium convolutum SP108	*Eggers, Chauveau & Heaton H09019* (ICN)	HQ606758	HQ606868	HQ607087
Sisyrinchium convolutum SP806	*Chauveau H09038* (UPS)	KF577269	KF577356	KF577195
Sisyrinchium convolutum SP818	*Chauveau H11021* (ICN)	KF577273	KF577360	KF577199
Sisyrinchium cuspidatum SP015	*RBGE 20040218* (living collection)	KF577211	KF577297	KF577136
Sisyrinchium cuspidatum SP038	*Eggers, Chauveau & Aubert H09011* (ICN)	HQ606725	HQ606835	HQ607054
Sisyrinchium demissum SP034	*Chauveau H09009* (ICN)	HQ606723	HQ606833	HQ607052
Sisyrinchium demissum SP182	*Chauveau s.n.* (ICN 190685)	KF577257	KF577343	KF577182
Sisyrinchium densiflorum SP119	*Eggers & Souza-Chies 321* (ICN)	HQ606760	HQ606870	HQ607089

Sisyrinchium elmeri SP813	*Chauveau & Heaton H09042* (ICN)	HQ606806	HQ606916	HQ607134
Sisyrinchium fasciculatum SP833	*Eggers & Souza-Chies 527* (ICN)	KF577275	KF577362	KF577201
Sisyrinchium fiebrigii SP121	*Eggers & Souza-Chies 325* (ICN)	HQ606761	HQ606871	HQ607090
Sisyrinchium foliosum SP072	*Eggers & Souza-Chies 262* (ICN)	KF577227	KF577313	KF577152
Sisyrinchium foliosum SP073	*Eggers & Souza-Chies 281* (ICN)	HQ606743	HQ606853	HQ607072
Sisyrinchium funereum SP035	*R. & E. Heaton SIS126.00* (living collection)	HQ606724	HQ606834	HQ607053
Sisyrinchium graminifolium SP183	*Eggers & Chauveau H09024* (ICN)	HQ606783	HQ606893	HQ607111
Sisyrinchium graminifolium SP800	*UPSBG SP800-01* (living collection)	KF577267	KF577354	KF577193
Sisyrinchium hasslerianum SP146	*Eggers & Souza-Chies 377* (ICN)	HQ606769	HQ606879	HQ607098
Sisyrinchium hoehnei SP135	*Eggers & Souza-Chies 355* (ICN)	HQ606766	HQ606876	HQ607095
Sisyrinchium hoehnei SP144	*Eggers & Souza-Chies 366* (ICN)	KF577247	KF577333	KF577172
Sisyrinchium hoehnei SP152	*Eggers & Souza-Chies 391* (ICN)	HQ606772	HQ606882	HQ607101
Sisyrinchium hoehnei SP157	*Eggers & Souza-Chies 400* (ICN)	KF577250	KF577336	KF577175
Sisyrinchium hoehnei SP896	*Eggers & Souza-Chies 618* (ICN)	KF577281	KF577369	KF577208
Sisyrinchium idahoense SP040	*R. & E. Heaton SIS183.00* (living collection)	HQ606726	HQ606836	HQ607055
Sisyrinchium idahoense var. *macounii* SP041	*R. & E. Heaton SIS180.01* (living collection)	HQ606727	HQ606837	HQ607056
Sisyrinchium iridifolium SP057	*Chauveau H11016* (ICN)	KF577222	KF577308	KF577147
Sisyrinchium jamesonii AA012	*Eggers et al. 012-2011* (ICN, CUZ)	KF153081	KF577287	KF577125
Sisyrinchium jamesonii AA017	*Eggers et al. 017-2011* (ICN, CUZ)	KF153083	KF577289	KF577127
Sisyrinchium jamesonii AA040	*Eggers et al. 040-2011 (ICN, CUZ)*	KF153086	KF577292	KF577130
Sisyrinchium jamesonii AA041	*Eggers et al. 041-2011* (ICN, CUZ)	KF153087	KF577293	KF577131
Sisyrinchium laxinervium AA039	*Eggers et al. 039-2011* (ICN, CUZ)	KF153085	KF577291	KF577129
Sisyrinchium laxinervium AA047	*Eggers et al. 047-2011* (ICN, CUZ)	KF153088	KF577294	KF577132
Sisyrinchium laxum SP049	*Chauveau & Aubert H09014* (ICN)	HQ606730	HQ606840	HQ607059
Sisyrinchium limarinum SP803	*Chauveau H09056* (ICN)	KF577268	KF577355	KF577194
Sisyrinchium littorale SP042	*Chauveau H09012* (ICN)	HQ606728	HQ606838	HQ607057
Sisyrinchium littorale SP171	*UPSBG 1295-01* (living collection)	KF577254	KF577340	KF577178
Sisyrinchium longipes SP181	*Chauveau H09023* (ICN)	HQ606782	HQ606892	HQ607110
Sisyrinchium luzula SP043	*Eggers & Souza-Chies 212* (ICN)	KF577217	KF577303	KF577142
Sisyrinchium luzula SP128	*Eggers & Souza-Chies 341* (ICN)	HQ606763	HQ606873	HQ607092
Sisyrinchium macrocarpum ssp. *macrocarpum SP045*	*BGCU C19890652* (living collection)	KF577218	KF577304	KF577143

Sisyrinchium macrocarpum ssp. *macrocarpum* SP047	*R. & E. Heaton* (living collection)	KF577219	KF577305	KF577144
Sisyrinchium macrocarpum ssp. *macrocarpum* SP106	*UPSBG 1295-01* (living collection)	KF577238	KF577324	KF577163
Sisyrinchium macrocarpum ssp. *laetum* SP235	*Chauveau H09031* (ICN)	HQ606796	HQ606906	HQ607124
Sisyrinchium macrocarpum ssp. *macrocarpum* SP046	*Chauveau H09013* (ICN)	HQ606729	HQ606839	HQ607058
Sisyrinchium macrophyllum SP050	*Chauveau & Heaton H09015* (UPS)	HQ606731	HQ606841	HQ607060
Sisyrinchium mandonii SP807	*Chauveau & Heaton H09039* (ICN)	HQ606803	HQ606913	HQ607131
Sisyrinchium marchio SP036	*R. & E. Heaton* (living collection)	KF577215	KF577301	KF577140
Sisyrinchium marchio SP134	*Eggers & Souza-Chies 354* (ICN)	KF577246	KF577332	KF577171
Sisyrinchium marchio SP161	*Eggers & Souza-Chies 407* (ICN)	HQ606775	HQ606885	HQ607104
Sisyrinchium marchio SP230	*Eggers & Souza-Chies 318* (ICN)	KF577263	KF577349	KF577188
Sisyrinchium marchioides SP229	*Eggers & Souza-Chies 319* (ICN)	KF577262	KF577348	KF577187
Sisyrinchium megapotamicum SP051	*Eggers & Souza-Chies 216* (ICN)	KF577220	KF577306	KF577145
Sisyrinchium megapotamicum SP052	*Eggers & Souza-Chies 236* (ICN)	HQ606732	HQ606842	HQ607061
Sisyrinchium miamense SP053	*R. & E. Heaton SIS176.00* (living collection)	HQ606733	HQ606843	HQ607062
Sisyrinchium micranthum SP054	*Eggers & Souza-Chies 242-A* (ICN)	HQ606734	HQ606844	HQ607063
Sisyrinchium micranthum SP055	*Eggers & Souza-Chies 210* (ICN)	KF577221	KF577307	KF577146
Sisyrinchium micranthum SP056	*Chauveau & Heaton H09054* (ICN)	HQ606735	HQ606845	HQ607064
Sisyrinchium micranthum SP058	*Eggers & Souza-Chies 261-H* (ICN)	KF577223	KF577309	KF577148
Sisyrinchium micranthum SP059	*Eggers & Souza-Chies 244-E* (ICN)	HQ606736	HQ606846	HQ607065
Sisyrinchium micranthum SP060	*Eggers & Souza-Chies 261-K* (ICN)	HQ606737	HQ606847	HQ607066
Sisyrinchium micranthum SP061	*Eggers & Souza-Chies 264-K* (ICN)	KF577224	KF577310	KF577149
Sisyrinchium micranthum SP062	*Eggers & Souza-Chies 234* (ICN)	HQ606738	HQ606848	HQ607067
Sisyrinchium micranthum SP185	*Eggers & Souza-Chies 261-F* (ICN)	KF577258	KF577344	KF577183
Sisyrinchium micranthum SP186	*Eggers & Souza-Chies 266-F* (ICN)	HQ606785	HQ606895	HQ607113
Sisyrinchium micranthum SP187	*Eggers & Souza-Chies 282-M* (ICN)	KF577259	KF577345	KF577184
Sisyrinchium micranthum SP188	*Eggers & Souza-Chies 233* (ICN)	KF577260	KF577346	KF577185
Sisyrinchium micranthum SP189	*Eggers & Souza-Chies 229-A* (ICN)	HQ606786	HQ606896	HQ607114
Sisyrinchium micranthum SP190	*Eggers & Souza-Chies 251* (ICN)	HQ606787	HQ606897	HQ607115
Sisyrinchium micranthum SP812	*R. & E. Heaton SIS106.02 (living collection)*	KF577270	KF577357	KF577196
Sisyrinchium micranthum SP839	*Chauveau & Aubert H09052* (ICN)	HQ606815	HQ606925	HQ607143

Sisyrinchium minus SP063	*Eggers & Souza-Chies 230* (ICN)	HQ606739	HQ606849	HQ607068
Sisyrinchium minus SP064	*Eggers 696 (ICN)*	KF577225	KF577311	KF577150
Sisyrinchium minutiflorum SP065	*Eggers & Souza-Chies 268* (ICN)	KF577226	KF577312	KF577151
Sisyrinchium minutiflorum SP066	*Eggers & Souza-Chies 285* (ICN)	HQ606740	HQ606850	HQ607069
Sisyrinchium montanum SP068	*R. & E. Heaton SIS114.02* (living collection)	HQ606741	HQ606851	HQ607070
Sisyrinchium nashii SP233	*Chauveau H09029* (ICN)	HQ606794	HQ606904	HQ607122
Sisyrinchium nidulare SP071	*Eggers & Souza-Chies 240* (ICN)	HQ606742	HQ606852	HQ607071
Sisyrinchium nidulare SP130	*Eggers & Souza-Chies 348* (ICN)	KF577245	KF577331	KF577170
Sisyrinchium ostenianum SP165	*Eggers & Souza-Chies 415* (ICN)	KF577252	KF577338	KF577177
Sisyrinchium ostenianum SP170	*Eggers & Souza-Chies 434* (ICN)	HQ606777	HQ606887	HQ607106
Sisyrinchium ostenianum SP193	*Eggers & Souza-Chies 422* (ICN)	KF577261	KF577347	KF577186
Sisyrinchium pachyrhizum SP832	*Eggers & Souza-Chies 515* (ICN)	KF577274	KF577361	KF577200
Sisyrinchium palmifolium SP118	*Eggers & Souza-Chies 320* (ICN)	KF577242	KF577328	KF577167
Sisyrinchium palmifolium ssp. *fuscoviride* SP076	*R. & E. Heaton SIS199.00* (living collection)	HQ606744	HQ606854	HQ607073
Sisyrinchium palmifolium ssp. *giganteum* SP077	*Eggers & Souza-Chies 215* (ICN)	KF577229	KF577315	KF577154
Sisyrinchium palmifolium ssp. *palmifolium* SP075	*Chauveau H11017* (ICN)	KF577228	KF577314	KF577153
Sisyrinchium palmifolium ssp. *palmifolium* SP113	*Eggers & Souza-Chies 255* (ICN)	KF577239	KF577325	KF577164
Sisyrinchium palmifolium ssp. *palmifolium* SP175	*Chauveau H09020* (ICN)	HQ606779	HQ606889	HQ607107
Sisyrinchium palustre AA008	*Eggers et al. 008-2011* (ICN, CUZ)	KF153080	KF577286	KF577124
Sisyrinchium papillosum SP249	*Chauveau H09037* (ICN)	HQ606802	HQ606912	HQ607130
Sisyrinchium parvifolium SP224	*Eggers & Souza-Chies 237-C* (ICN)	HQ606791	HQ606901	HQ607119
Sisyrinchium parvifolium SP251	*Eggers & Souza-Chies 464* (ICN)	KF577265	KF577351	KF577190
Sisyrinchium patagonicum SP037	*H09010* [Parc Botanique de Launay, Université Paris-Sud 11, France]	KF577216	KF577302	KF577141
Sisyrinchium patagonicum SP177	*MNHN 19902* (living collection)	KF577256	KF577342	KF577181
Sisyrinchium patagonicum SP223	*Chauveau H09027* (ICN)	HQ606790	HQ606900	HQ607118
Sisyrinchium pearcei SP180	*Chauveau H09022* (ICN)	HQ606781	HQ606891	HQ607109
Sisyrinchium platense SP083	*R. & E. Heaton* (living collection)	KF577230	KF577316	KF577155
Sisyrinchium platense SP084	*Eggers & Souza-Chies 187* (ICN)	HQ606746	HQ606856	HQ607075

Sisyrinchium platense SP834	*Eggers & Souza-Chies 530* (ICN)	KF577276	KF577363	KF577202
Sisyrinchium prealtum AA015	*Eggers et al. 015-2011* (ICN, CUZ)	KF153082	KF577288	KF577126
Sisyrinchium pruinosum SP816	*Eggers, Chauveau & Heaton H09048* (ICN)	HQ606807	HQ606917	HQ607135
Sisyrinchium purpurellum ssp. *trichospathum* SP132	*Eggers & Souza-Chies 350* (ICN)	HQ606764	HQ606874	HQ607093
Sisyrinchium purpurellum ssp.*purpurellum* SP151	*Eggers & Souza-Chies 387* (ICN)	KF577249	KF577335	KF577174
Sisyrinchium pusillum AA001	*Eggers et al. 001-2011* (ICN, CUZ)	KF153077	KF577283	KF577121
Sisyrinchium rambonis SP169	*Eggers & Souza-Chies 291* (ICN)	KF577253	KF577339	
Sisyrinchium rambonis SP835	*Eggers & Souza-Chies 571* (ICN)	HQ606812	HQ606922	HQ607140
Sisyrinchium rectilineum SP114	*Eggers & Souza-Chies 284* (ICN)	KF577240	KF577326	KF577165
Sisyrinchium rectilineum SP123	*Eggers & Souza-Chies 332* (ICN)	HQ606762	HQ606872	HQ607091
Sisyrinchium restioides SP096	*Eggers & Souza-Chies 217* (ICN)	KF577235	KF577321	KF577160
Sisyrinchium restioides SP097	*Eggers & Souza-Chies 252* (ICN)	HQ606754	HQ606864	HQ607083
Sisyrinchium rosulatum SP085	*ICN 187084* (ICN)	HQ606747	HQ606857	HQ607076
Sisyrinchium sarmentosum SP086	*Chauveau H09017* (ICN)	HQ606748	HQ606858	HQ607077
Sisyrinchium scabrum SP088	*R. & E. Heaton SIS114.02* (living collection)	HQ606750	HQ606860	HQ607079
Sisyrinchium scariosum SP087	*Eggers & Souza-Chies 277* (ICN)	HQ606749	HQ606859	HQ607078
Sisyrinchium scariosum SP089	*PABG IRI117*(living collection)	KF577231	KF577317	KF577156
Sisyrinchium sellowianum SP090	*Eggers & Souza-Chies 238* (ICN)	KF577232	KF577318	KF577157
Sisyrinchium sellowianum SP091	*Eggers & Souza-Chies 253* (ICN)	HQ606751	HQ606861	HQ607080
Sisyrinchium setaceum SP092	*Eggers & Souza-Chies 214* (ICN)	HQ606752	HQ606862	HQ607081
Sisyrinchium setaceum SP093	*Eggers & Souza-Chies 225* (ICN)	KF577233	KF577319	KF577158
Sisyrinchium soboliferum SP148	*Eggers & Souza-Chies 381* (ICN)	HQ606771	HQ606881	HQ607100
Sisyrinchium soboliferum SP117	*Eggers & Souza-Chies 317* (ICN)	KF577241	KF577327	KF577166
Sisyrinchium striatum SP094	*UPSBG R07130* (living collection)	KF577234	KF577320	KF577159
Sisyrinchium striatum SP095	*Chauveau H09018* (ICN)	HQ606753	HQ606863	HQ607082
Sisyrinchium striatum SP844	*Chauveau H11008* (ICN)		KF577366	KF577205
Sisyrinchium tenuifolium SP184	*Chauveau & Pautz H09025* (ICN)	HQ606784	HQ606894	HQ607112
Sisyrinchium tinctorium SP240	*Chauveau & Heaton H09034* (ICN)	HQ606801	HQ606911	HQ607129
Sisyrinchium tinctorium SP814	*R. & E. Heaton SIS143.06* (living collection)	KF577271	KF577358	KF577197
Sisyrinchium tinctorium SP815	*R. & E. Heaton SIS143.02* (living collection)	KF577272	KF577359	KF577198
Sisyrinchium uliginosum SP124	*Eggers & Souza-Chies 333* (ICN)	KF577243	KF577329	KF577168

Sisyrinchium uliginosum SP126	*Eggers & Souza-Chies 339* (ICN)	KF577244	KF577330	KF577169
Sisyrinchium uliginosum SP154	*Eggers & Souza-Chies 393* (ICN)	HQ606773	HQ606883	HQ607102
Sisyrinchiumvaginatum ssp. *vaginatum* SP231	*Eggers & Souza-Chies 374* (ICN)	KF577264	KF577350	KF577189
Sisyrinchium vaginatum ssp. *vaginatum* SP226	*Eggers & Souza-Chies 263* (ICN)	HQ606792	HQ606902	HQ607120
Sisyrinchium vaginatum ssp.*vaginatum* SP836	*Eggers & Souza-Chies 463A* (ICN)	KF577277	KF577364	KF577203
Sisyrinchium weirii SP099	*Eggers & Souza-Chies 248* (ICN)	HQ606755	HQ606865	HQ607084
Sisyrinchium wettsteinii SP115	*Eggers & Souza-Chies 250* (ICN)	HQ606759	HQ606869	HQ607088
Solenomelus pedunculatus SP221	*Chauveau H09044* (ICN)	HQ606789	HQ606899	HQ607117
Solenomelus segethi SP005	*Chauveau & Aubert H09001* (ICN)	HQ606706	HQ606816	HQ607036

Abbreviations: PABG, Botanical Garden of Porto Alegre (Brazil); MNHN, Botanical Garden of the Muséum National d'Histoire Naturelle (Paris, France); MNHN, Botanical Garden of the Muséum National d'Histoire Naturelle (Paris, France); BGCU, Botanical garden - Cambridge Universiy (United Kingdom); UPSBG, Botanical Garden of the University Paris-Sud (France); RBGE, Royal Botanical Garden of Edinburgh (United Kingdom).

Table S2. Primers used for amplification and sequencing in this study.

Barcode region	Name (direction)	Primer sequence (5'-3')	Source
matK	matK-Xf (forward)	TAATTTACGATCAATTCATTC	http://www.kew.org/barcoding/update.html
	matK-3.2r (reverse)	CTTCCTCTGTAAAGAATTC	http://www.kew.org/barcoding/protocols.html
trnH-psbA	psbA (forward)	GTTATGCATGAACGTAATGCTC	(Shaw *et al*. 2005)
	trnH^{GUG} (reverse)	CGCGCATGGTGGATTCACAATCC	(Shaw *et al*. 2005)
ITS	ITS5b (forward)	GGAAGTAAAAGTCGTAACAAG	Modified from ITS5 (White *et al*., 1990)
	ITS4 (reverse)	TCCTCCGCTTATTGATATGC	(White *et al*., 1990)
	ITS-38f (forward)	CTGCGGAAGGATCATTGTC	(Chauveau *et al*. 2011)

References

Chauveau O, Eggers L, Raquin C, Silvério A, Brown S, Couloux A, Cruaud C, Kaltchuk-Santos E, Yockteng R, Souza-Chies TT, Nadot S (2011) Evolution of oil producing trichomes in *Sisyrinchium* (Iridaceae): insights from the first comprehensive phylogenetic analysis of the genus. *Annals of Botany*, **107**, 1287–1312.

http://www.kew.org/barcoding/protocols.html

http://www.kew.org/barcoding/update.html

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, editor. PCR protocols: a guide to methods and applications. San Diego, CA: Academy Press, p. 315-322.

Table S3: Sample frequencies per species of *Sisyrinchium* and relatives.

Number of samples per species.	Frequency	
	ITS presente	ITS absent
One	53	52
Two	27	28
Three	11	11
Four	3	3
Five	2	2
Six	1	1
16	1	1
N	98	98

Capítulo 2

Applying DNA barcoding to South-American Tigridieae (Iridaceae): restricted variation limits identification in related species

Artigo em preparação para submissão à revista *Molecular Ecology Resources*

Applying DNA barcoding to South-American Tigridieae (Iridaceae): restricted variation limits identification in related species

TIAGO LUIZ DA SILVA ALVES, OLIVIER CHAUVEAU, EUDES MARIA STIEHL-ALVES, LEANDRO DAL RI, LILIAN EGGERS and TATIANA TEIXEIRA DE SOUZA-CHIES*

Programa de Pós-Graduação em Botânica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, Av. Bento Gonçalves, 9500 - Prédio 43433, Brazil

Keywords: *rbcL, matK*, *trnH-psbA*, *Calydorea*, *Cypella*, *Gelasine* , *Herbertia, Phallocalis*.

*Corresponding author: Tatiana Teixeira de Souza Chies, Fax: 51 3308 7755; e-mail: tatiana.chies@ufrgs.br

Abstract

The main objective of DNA barcoding is identifying species accurately using short, polymorphic, standardized and universal DNA sequences. DNA barcoding methods were tested among up to 17 species distributed into seven New World genera from tribe Tigridieae (Iridaceae: Iridoideae). Sampling was geographically concentrated and varies from 65 to 58 individuals depending on the barcode region in question. The samples analyzed in this study belong to putatively related species according to molecular and morphological evidences. We tested a set of non-combined markers including plastid *rbcL*, *matK*, *trnH-psbA* and nuclear ITS. Variability parameters, evaluation of the barcoding gap, TAXONDNA similarity-based analyzes and three different tree-based methods were used to assess sequence variation and identification performances. The only barcoding region that presented significant low universality was ITS (39.06%). The levels of species identification obtained from plastid barcodes were low, especially for *rbcL*. The *trnH-psbA* provided slightly better results than *rbcL*, but the best identification performance was achieved by *matK*, with 50% of the species and up to 68.18% of the individuals properly identified. Plastid *matK* is the most recommended region for DNA barcoding among Tigridieae species up to now. In spite of some limitations, different research areas such as ecology and taxonomy could use DNA barcoding tools to achieve species identification. It is especially useful among these Tigridieae species, since these plants are very homogeneous morphologically when in the absence of flowers, being virtually indistinguishable without using molecular tools.

Introduction

The use of DNA barcoding as universal tool for species identification accomplished ten years in 2013. The proposal of using short, variable, standardized and universal DNA sequences to identify species (CBOL Plant Working Group, 2009; Hebert *et al*. 2003; Hollingsworth *et al*., 2009; Savolainen *et al*. 2005; Tautz *et al*., 2003) is the essential aim of the DNA barcoding. Since the pioneer work of Herbert *et al*. (2003), the establishing of such approach has gained widespread application in a wide array of organisms. In spite of the undeniable general progress and achievements of DNA barcoding, discriminating among animal species has reached more success than among plants (Fazekas *et al*., 2009). The mitochondrial gene of cytochrome *c* oxidase subunit 1 (*cox1/*CO1) accomplishes well the central requirements of DNA barcoding in animals. However, unlike animals, the use of multiple DNA barcoding markers is required to attain effective species identification in plants (CBOL Plant Working Group, 2009; Hollingsworth *et al*., 2009). Several *loci* have already been tested in single or multiple *loci* combination, usually providing lower resolution ability when compared to that obtained for animals using CO1 (Alves *et al*., in press; Fazekas *et al*., 2008; Fazekas *et al*., 2009; Lahaye *et al*. 2008; Maia *et al*., 2012; Newmaster *et al*. 2008; Roy *et al*., 2010; Seberg & Petersen, 2009; Spooner, 2009). In this way, it is consensual that a single *locus* is not enough to provide successful identification among all plant species. The sequences used for DNA barcoding in plants are mainly located on the plastid genome (Hollingsworth *et al*., 2009). The use of few plastid sequences is becoming gradually disseminated, especially markers such as *rbcL*, *matK* and the intergenic spacer *trnH-psbA* (CBOL Plant Working Group, 2009). The nrITS is the only region outside plastid genome that has widespread use as DNA barcoding marker (CBOL Plant Working Group, 2009; Hollingsworth *et al*., 2009).

Choosing a suitable DNA barcoding marker is certainly a complicated task since it is hard to establish a balance between two central requirements of DNA barcoding studies: universality and resolution power. It is quite challenging to find markers with enough level of variation for discriminating species along with conservative flanks for universal primers. Another limiting is the priority of using multicopy *loci* as preferential for DNA barcoding studies because they greatly facilitate the PCR amplification process. Finally, the specific use of nuclear ITS has some inherent difficulties to deal with, since nuclear markers are susceptible to fungal cross-amplification, presence of paralogous copies and difficulties in sequencing and amplifying procedures (Hollingsworth, 2011).

The evolutionary history of plants is relevant to understand why plants are usually more difficult to be classified into properly established species. Plant species boundaries are very influenced by polyploidy, natural hybridization and apomixis (Fazekas *et al*., 2009). DNA barcoding performance is considered poor in groups that have experimented either hybridization or introgression (Alves *et al.,* in press; Roy *et al*., 2010; Spooner, 2009). Some estimates refer that about 15% of the angiosperms species evolves through polyploidization events (Wood *et al*., 2009). DNA barcoding is not exclusively based on phylogenetic centered approaches, but the use of phylogenetic signal is widespread in DNA barcoding as basis for tree-based analysis to test species monophyly as a criterion of identification (Alves *et al.* in press; Theodoridis *et al*., 2012; Zhang *et al*., 2012). The retention of ancestral polymorphisms in the presence of incomplete lineage sorting in plants is quoted as another major source of confusion, which can preclude phylogenetic inference (Maddison & Knowles, 2006; Yu *et al*., 2011b) with consequences in DNA barcoding identifications. The abovementioned phenomena characterize evolutionary histories that are better described by networks than by phylogenetic trees (Huson and Bryant, 2006). Indeed,

any form of reticulation is a potential source of phylogenetic incongruence that certainly influences the DNA barcoding performance since the quality of the phylogenetic signal is generally associated with better DNA barcoding performances.

Most of the DNA barcoding approaches were designed to inventory plants restricted to a particular area, usually biodiversity hotspots like tropical forests (Wang *et al*., 2010). The use of DNA barcoding to identify members of a restricted geographic area commonly achieves more accuracy than studies focused on natural groups, such as a genera or family. The inherent larger divergence exhibited among distantly related plants in local DNA barcoding works favors species identification. A more challenging situation is observed when we are dealing with closely related taxa because sequences tend to be more homogeneous, precluding species assignment. It is common that taxonomic centered studies present low to moderate identification performances (Alves *et al*., in press; Maia *et al*., 2012; Pang *et al.*, 2010; Roy *et al.*, 2010; Seberg & Petersen, 2009; Spooner, 2009; Starr *et al.*, 2009).

The tribe Tigridieae (Iridaceae: Iridoideae) is a New World exclusive group of herbaceous monocotyledons with the main diversity centers on Andean South America and Mexico (Rodriguez & Sytsma, 2006). Tigridieae is classified into two subtribes called Cipurineae and Tigridiineae (Goldblatt, 1990). Rodriguez & Sytsma (2006) suggested a bimodal geographic distribution for the two recognized subtribes of Tigridieae, stating that Cipurineae has a center of diversification in South America, whereas Tigridiineae is predominantly distributed in Mexico and Guatemala.

The tribe Tigridieae is characterized by the typical presence of bulbs, and the majority of its members bear flowers remarkably conspicuous and ornamental, presenting a wide range of different colors and shapes (Rodriguez & Sytsma, 2006). The floral

morphology is essential to identify members of Tigridieae because most of these plants are virtually indistinguishable when in vegetative form due to the extreme homogeneity of vegetative characters. Furthermore, the flowering period of these plants is short, the flowers are labile and very hard to preserve in exsiccates (Rodriguez & Sytsma, 2006). The difficulty of assigning species identification in the absence of reproductive structures and the fragility of the flowers make this group an interesting target for DNA barcoding approaches. Previous phylogenetic studies focused on Tigridieae evidenced two major clades, confirming partly the previous subdivisions into the subtribes Tigridiineae and Cipurineae (Chauveau *et al.*, 2012; Rodriguez & Sytsma, 2006). Alternatively, Chauveau *et al*. (2012) subdivided the tribe into two main clades named clade A (only Cipurineae species) and clade B (most Tigridiineae species). Tigridieae accounts with several genera with unclear delimitations, such as *Cypella* Herb., *Calydorea* Herb. and *Herbertia* Sweet in Clade A; and *Tigridia* Juss. and *Gelasine* Herb. in clade B (Rodriguez & Sytsma, 2006; Chauveau *et al.*, 2012). The pollination systems are implicated in the diversification of Tigridieae and it seems that floral variation in the tribe is driven by adaptations to pollinators (Chauveau *et al.*, 2012).

This work aims to apply DNA barcoding approach using the CBOL most recommended *loci* for plants (*rbcL*, *matK*, *trnH-psbA* and ITS) in tribe Tigridieae in order to achieve reliable species identification. Phylogenetic methods (tree-based analyses) and sequence similarity-based methods were used to access species identification in Tigridieae focused primarily on the clade A. We evaluated (i) the universality of the tested *loci* (ii), the *barcoding gap* as estimative of intra- versus interspecific variation and (iii) the accuracy in identifying species.

Material and methods

Taxon sampling

The sampling covered geographically distributed species within the South American biodiversity center of Tigridieae, (Rodriguez & Sytsma, 2006). Our sample covered a total of 17 species distributed into seven genera as follows: *Herbertia* (six species), *Cypella* (three species), *Cipura* Aubl. (one species), *Calydorea* (three species), *Kelissa* Ravenna (one species), *Gelasine* (two species) and *Phallocalis* Herb.(one species). The majority of these samples were from species nested in Clade A (Chauveau *et al*., 2012), but samples from two poorly undestood genera from Clade B, *Gelasine* and *Phallocalis*, were sampled as well. Goldblatt (1990) stated that Tigridieae accounts with 18 genera, and threfore our sample comprised about 39% of the generic diversity of Tigridieae. The total number of individuals sampled varies from 58 to 65 according to the DNA barcoding region (Table 1). All vouchers were deposited in the herbarium of the Universidade Federal do Rio Grande do Sul (ICN).

DNA extraction, amplification and sequencing

Total genomic DNA was extracted from leaves dried on silica-gel using the cetyl trimethylammonium bromide (CTAB) protocol adapted from Doyle & Doyle, (1987). Polymerase chain reactions (PCR) were prepared in 25 µL reaction mixes containing 20 to 100 ng of genomic DNA; 1 µM of both reverse and forward primers; 100 µM of dNTP; *Taq* polymerase buffer (1X); 1.5 mM of $MgCl_2$ and 1.5 U of Taq DNA Polimerase (Invitrogen,

São Paulo, SP, Brazil). We also used 1 to 2.5 µL of pure DMSO for *trnH-psbA* and ITS PCR procedures in order to achieve better amplification success. PCR amplifications were performed in a thermocycler Veriti 96 Well Thermal Cycle (Applied Biosystems). Visualization of the DNA fragments on a 1% agarose gel were used to verify the efficiency of PCR reactions. The primer sequences and thermocycling conditions are specified in Table S1. Reverse and forward primers were added to the unpurified PCR products and they were dried using a vacuum centrifuge and subsequently sequenced by Cogenics (Meylan, France) in an AB1 PRISM authomatic sequencer. The *matK* sequences had more than 1kb and two extra primers were used for sequencing (Table S1).

Data analyses

CODONCODE ALIGNER 4.2.2 (CodonCode Corp., Dedham, MA, USA) was used to assemble raw bidirectional sequences. The resulting contigs were visually inspected using MEGA v.5.2.2 (Tamura *et al*., 2011) and aligned with MUSCLE (Edgar, 2004). Pairwise K2P (Kimura 2-parameter) distances for every DNA regions were calculated using TAXONDNA (Meier *et al*., 2006) to evaluate intraspecific and interspecific divergence. The relative distribution of the frequencies of K2P pairwise genetic distances were performed in order to estimate the barcoding gap.

Two identification tools from TAXONDNA program and three tree-based methods were used to assess species identification. The best match (BM) and the best close match (BCM) modes of TAXONDNA were conducted using the K2P distance as model. The program output yielded the proportion of individuals identified according to BM and BCM methods. Neighbor joining (NJ), maximum parsimony (MP) and weighted pair group

method with arithmetic mean (UPGMA) were performed on MEGA v.5.2.2 (Tamura *et al*., 2011). All resulting trees were evaluated using 1000 bootstrap iterations as node support values for the branches. Only monophyletic clusters which nodes supported by at least 60% of bootstrap were considered correctly identified.

Results

Universality of PCR amplification and DNA sequencing

Among the four tested DNA barcoding markers, the only that presented appropriate universality was *rbcL*, with 95.38% of success (Table 2). In spite of the 100% of samples amplified, the total universality obtained for *trnH-psbA* reached 84.21% due to sequencing inefficiency observed for all samples (10) of *Cypella*. The reasons of this unexpected taxon-specific failure remain in debate (see discussion). The universality of *matK* was lower (68.75%) mostly due to sequencing failures. Lastly, ITS spacers showed a relatively satisfactory PCR success, but this marker presented a universality of only 39.06 due to significant failures on sequencing. Furthermore, even high quality reads from such marker were not easy to align after assembling, requiring significant manual editing, principally concerning the indels. Thus, we disregarded the use of ITS in the identification analyses due to its remarkable low universality. The remaining markers where analyzed independently. We also avoided combinations among DNA barcoding markers for the identification analyses due to incongruences that resulted in poorly sampled matrices, overestimating the barcoding gap and consequently the success rate of identification.

Table 1: Sampling and universality (PCR and sequencing success ratio) from three plastidal and nrITS barcode markers.

Barcode region	Tested samples, *N*	Amplified samples. *N*	Sequenced samples, *N*	Universality, %
rbcL	65	62	62	95.38
matK	64	57	44	68.75
trnH-psbA	58	58	48	84.21
ITS	64	54	25	39.06

Table 2. Divergence metrics and alignment length of the three DNA barcoding markers retained for analyses.

Barcode region	Alignment length (bp)	Variable characters, N (%)	Parsimony informative characters, N (%)	Overall K2P distance, %
rbcL	553	30 (5.42)	19 (3.44)	0.6
matK	1017	108 (10.62)	66 (6.49)	1.3
trnH-psbA	688	33 (4.80)	23 (3.34)	0.6

Variability levels and barcoding gap

The number of variables characters (VC), parsimony informative characters (PIC) and the overall K2P distance were very low for all retained markers (Table 2). Our results showed that the number of VC and PIC reached a maximum of just 10.62% for *matK*. The values of these metrics were around 5% for both *rbcL* and *trnH-psbA*. Conversely, the overall K2P distance for *matK* was only 1.3% and it was even lower for *rbcL* and *trnH-*psbA, with 0.6% for both sequences.

The results obtained with the histograms of K2P frequency distances were highly homogeneous among DNA barcoding markers (Figure 1). There is no barcoding gap since

the distribution of intra- and interspecific distances are highly overlapped in all histograms. The magnitude of interspecific distances was extremely low with values never exceeding 3.5%.

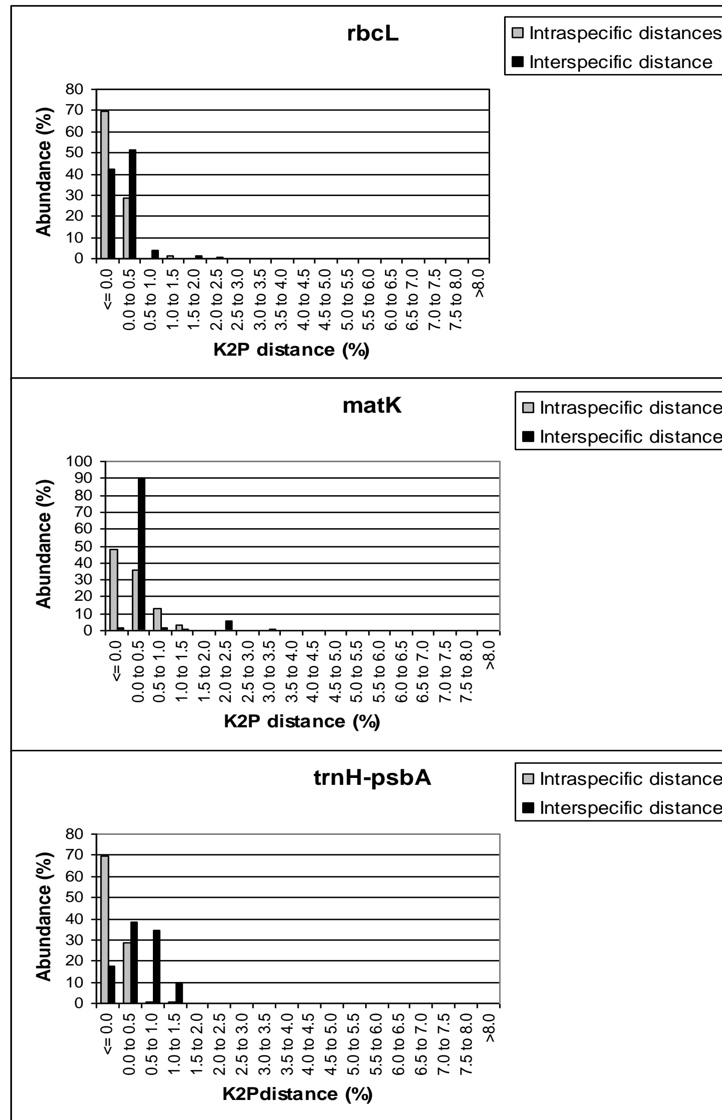


Figure 1. Intra- and interspecific distribution of the K2P distances for the three DNA barcoding markers retained for analyses.

Species identification

Species discrimination was calculated in terms of individuals identified using the five identification methods (Table 3). Moreover, the number of species successfully identified was depicted for the three tree-based methods (Table 4). The use of the *rbcL* region identified the lower number of samples (6.45 to 8.06%). Resolving power for the *matK* region was the higher one and varied between 34.09% and 68.18%. The *trnH-psbA* presented intermediary performance with correct identifications varying between 16.67% and 50.00%.

The proportion of correct assignments did not differ between BM and BCM methods, but the identification power of such methods were considerably higher than that achieved by tree-based methods, except for *rbcL*, which presented slightly better performance using the tree-based methods.

Barcode region	*N* samples	Correct identification, %						
		NJ	MP	UPGMA	BM	BCM	Mean	s.d.
rbcL	62	8.06	8.06	6.45	6.45	6.45	7.09	0.39
matK	44	34.09	34.09	34.09	68.18	65.91	47.27	8.08
trnH-psbA	48	22.92	16.67	22.92	50.00	50.00	32.50	7.23

Table 3. Percentage of samples accurately identified, mean of identification and standard deviation (s.d.) using five different methods for the three DNA barcoding markers retained for analyses.

When we consider the number of species identified, the results are slightly different because the number of conspecific samples (Table 3 and 4) influences the success rate of identification. Conversely, the results indicated a better performance for *matK* (50.00% of species identified) followed by *trnH-psbA* and *rbcL* with 50.00% to 35.71% and 17.65% to11.76% of species identified respectively. Figure 2 provide a visual perspective of the resulting clusters and branch node supports (collapsed over 60% of bootstrap) for *matK,* the best overall results using a NJ as method.

		Correct identification, %		
Barcode region	*N* species	NJ	MP	UPGMA
rbcL	17	17.65	17.65	11.76
matK	14	50.00	50.00	50.00
trnH-psbA	14	35.71	50.00	28.57

Table 4. Percentage of species accurately identified using three different tree-based methods for the three DNA barcoding markers retained for analyses.

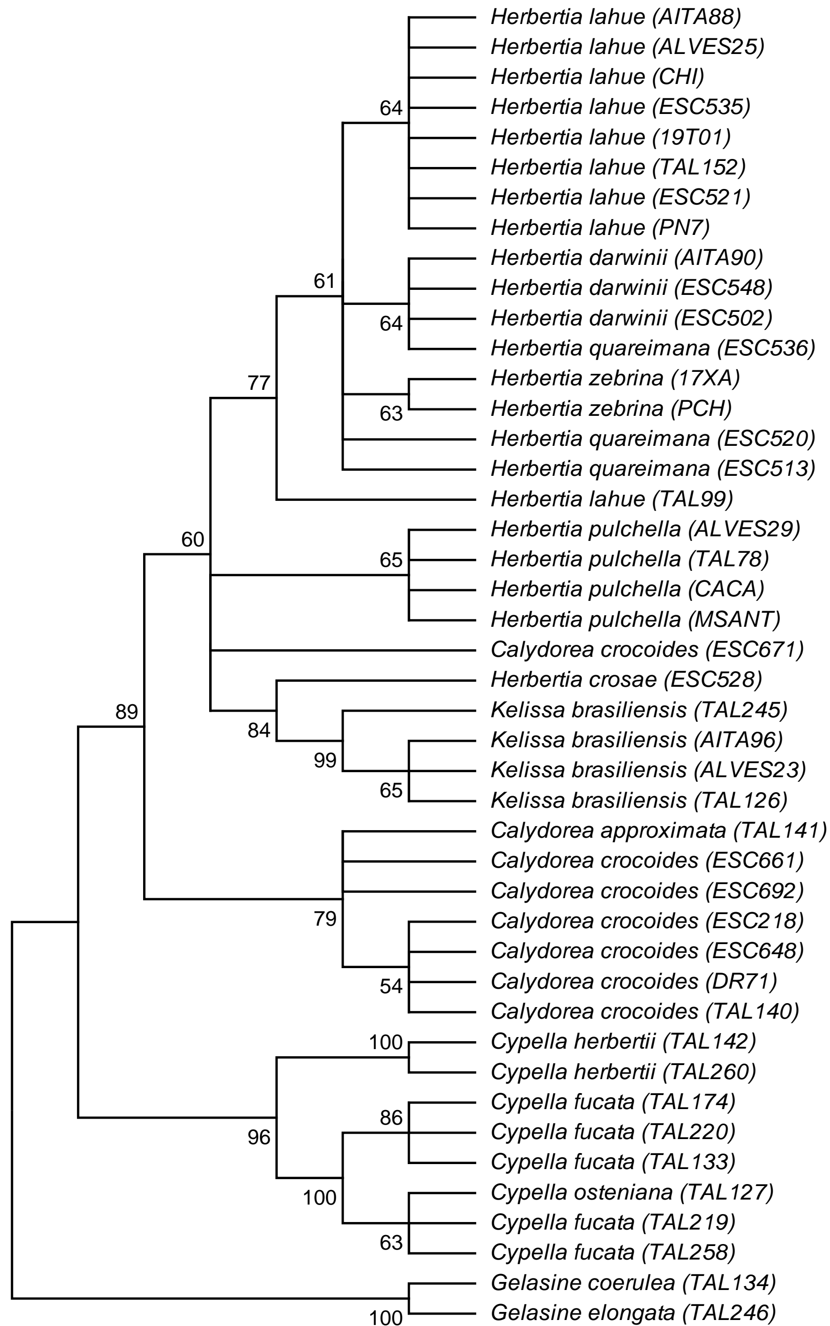


Figure 2. Phylogenetic tree based on NJ method of 48 Tigridieae samples using *matK locus*. Unrooted treess were performes in MEGA v.5.2.2 (Tamura *et al.*, 2011) using K2P as model of substitution under 1,000 bootstrap iterations and branches collapsed over 50% of node support.

Discussion

Universality levels among markers

High universality is one of the main requirements for accurate DNA barcoding approaches (CBOL Plant Working Group, 2009; Hollingsworth *et al*., 2009). In this way, the *rbcL* region is recognized by its high universality when compared to other DNA barcoding *loci* (Hollingsworth *et al*., 2009), and our results confirm such tendency. The *matK* reached limited universality in this study, presenting both amplification and sequencing problems, in spite of the use of two set of primers for sequencing. The size of more than 1 kb may be the reason for this lower universality since larger sequences are more susceptible to failures in amplifying and sequencing procedures. The *matK* region often exhibits low amplification and sequencing rates due to the low universality of the available primers, causing difficulties to obtain fully bidirectional sequences (Fazekas *et al*. 2008; Yu *et al*., 2011a).

The non-coding region *trnH-psbA* is a fairly variable intergenic spacer and such evolutionary rate allows its use to amplify very divergent plant lineages (Shaw *et al*., 2005, 2007). This intergenic spacer was considered easy to amplify and align in studies at generic level (Lahaye *et al*. 2008; Ren *et al*. 2010). Nevertheless, alignment ambiguities related to frequent micro inversions are frequent in *trnH-psbA* (Sang *et al*., 1997; Whitlock *et al*., 2010), impairing the quality of the sequencing (Fazekas *et al*., 2008). Our results showed that *trnH-psbA* was easily amplified across all samples, but sequencing fails specifically when samples from *Cypella* were tested. The reasons concerning this unexpected taxon-

specific failure are still in investigation, but it might be related to the presence of insertions and consequently longer sequence lengths of exclusive occurrence within *Cypella* species (data not shown). Nuclear ITS is increasingly used for DNA barcoding due to its strong identification ability (Li *et al*., 2011), but possess specific amplification disadvantages (Hollingsworth, 2011). The ITS universality ratios are usually above 90% (*eg*., Li *et al*., 2012; Yang *et al*., 2012; Zhang *et al*., 2012), demonstrating that the low levels of universality exhibited by ITS in Tigridieae in this study are unusual. Specific failures on ITS sequencing may be related to the presence of paralogous ITS copies (Hollingsworth, 2011; Nieto Feliner & Roselló, 2007). DNA barcoding methods unfortunately can not identify the occurrence of different genomic copies when concerted evolution is not fully operating. Cloning may be the better solution to distinguish among putative paralogous copies, but cloning methods are certainly out of the range of routine DNA barcoding studies. Further researches focused on ITS characterization in Tigridieae are necessary in order to test whether paralogy is present or not. The inherent features of ITS in Tigridieae put in doubt the usefulness of this *locus* as an effective DNA barcoding marker for this tribe up to now.

Overall divergence and evaluation of the barcoding gap

Our results showed that both intra- and interspecific divergences were very low (Figure 1), causing fully overlapped distributions in the K2P distance histograms among all markers analyzed. The occurrence of strongly overlapped distance distributions is common among plant DNA barcoding studies (Lahaye *et al*., 2008). In fact, barcoding gap based analysis seems to be more successful in identifying species among animals, but not among

land plants (Ashfaq *et al*., 2013). It was reported that plastid *loci* often do not exhibit distinctive barcoding gap, particularly when closely related species are examined (Alves *et al*., in press; Lahaye *et al*., 2008). The absence of any barcoding gap, the reduced overall K2P distance and the low number of PICs and VCs observed in Tigridieae suggest that every marker lacks polymorphism, indicating that the species analyzed possess very homogeneous plastid sequences.

At first glance, it seems peculiar that samples belonging to seven different genera exhibit very low polymorphism. Nevertheless, there are evidences that our sample comprises closely related species. Although belonging to distinct genera, the majority of the species analyzed shares the same origin in the Clade A (only Cipurineae species) according to the phylogeny proposed by Chauveau *et al*., (2012). These plants are also morphologically homogeneous, with most striking differences occurring in the floral shape, a trait probably influenced by adaptations to pollinators (Chauveau *et al*., 2012; Rodriguez & Sytsma, 2006). Genera such as *Cypella, Calydorea* and *Herbertia* are typically characterized by the lack of obvious distinctive floral characters (Chauveau *et al*., 2012). Furthermore, *Herbertia*, *Kelissa* and *Cypella* exhibit strong morphological homogeneity when in vegetative stage (Chauveau *et al*., 2012). These set of evidences suggests that the majority of the species in this study are closely related *taxa*, and it is in congruence with the low interspecific variation depicted in all distribution histograms. It seems that genetic distances within clades are prone to be smaller than those found among species belonging to different clades. Thus, our results indicate that the absence of the barcoding gap observed among all plastid data is related to the natural proximity of these lineages within Tigridieae.

Identification ability

The low success of identification observed for *rbcL* and *trnH-psbA* intergenic spacer are in agreement with the low polymorphism exhibited by such *loci* (Table 1). It is not surprisingly this failure of *rbcL* to separate species, since this region is recognized by its low identification rates (Ashfaq *et al*., 2013; Hollingsworth *et al*., 2009; Ren *et al*., 2010). A more polymorphic *trnH-psbA* usually performs better than *rbcL* and commonly works with relative success across several plant groups (Gonzalez *et al*., 2009; Kress *et al*., 2009; Kress *et al*., 2010). However, the *trnH-psbA* spacer showed very poor discrimination capacity in a comprehensive sampling from Sisyrinchieae focused on genus *Sisyrinchium* L. (Iridaceae) (Alves *et al*., in press). Our results indicate that *trnH-psbA* region identification success was relatively higher than that achieved for *rbcL*, but the overall identification ability remains modest. The results present in this study suggest that *trnH-psbA* region exhibits limited performance to distinguish among lineages within Tigridieae.

Among all plastid *loci* tested, *matK* presented the best identification capacity, reaching a maximum identification performance of 68.18% (BCM) and a half of the species correctly identified in all tree-based method. The *matK* region is the larger and most polymorphic sequence in our dataset (Table 1), and it is congruent with the best identification power observed (Tables 3 and 4). A very successful approach using *matK* was accomplished for Indian *Paphiopedilum* Pfitzer (Orchidaceae), with all species correctly identified (Parveen *et al*., 2012). As opposite, *matK* performed very poorly in *Ficus* L. (Moraceae) (Li *et al*., 2012) and in also in *Sisyrinchium* (Alves *et al*., in press), showing that *matK* capacity of identification varies significantly among different taxonomic

groups. It is possible that the lower number of samples (Table 2) have slightly influenced the identification power of *matK* by overestimating the distances. However, mainly due to the impediments of using ITS, *matK* really seems be the most appropriate marker to distinguish among the species analyzed since it presented more polymorphism and consequently the best identification power.

Conclusions

The limited capacity in identifying species using individual markers verified here are in congruence with other DNA barcoding researches focused on Iridaceae (Alves *et al*., is press; Seberg & Petersen, 2009). Among the tested markers, *matK* most closely met the requirements to distinguish among species, although it presented limited universality in this study. The best universality reached by *rbcL* and *trnH-psbA* were accompanied by the lack of divergence and low ability of species identification. Therefore, they are of restricted value as suitable DNA barcoding markers for these Tigridieae species. ITS region did not allow species discrimination among Tigridieae. The occurrence of incomplete concerted evolution leading to paralogy is the main hypothesis to explain ITS sequencing failures. Thus, the reliability of using this *locus* for DNA barcoding in Tigridieae remains doubtful. There are good evidences that DNA barcoding may not work in evolutionary complex lineages of plants (Alves *et al*., in press; Roy *et al*., 2010; Spooner, 2009). It seems to be the case of the majority of the genera analyzed here since these plants are very similar in vegetative habit, with shifts predominantly located on flower shape, a character remarkably influenced by adaptation to pollinators (Chauveau *et al*., 2012). In spite of some

universality restraints, the use of *matK* is promising to accomplish reliable identification once the universality problems are solved.

Species identification in Tigridieae using morphological traits exhibits a set of obstacles. The liability of the flowers in field, the fragility of reproductive structures in herbaria, the short blooming period and the vegetative homogeneity are among the most striking difficulties, which DNA barcoding approaches can overcome. At least three species analyzed in this work, *Herbertia zebrina* Deble (Deble, 2010), *Kelissa brasiliensis* (Baker) Ravenna and *Calydorea crocoides* (Iganci *et al*., 2011), present restricted geographic distribution, and our results may be helpful in conservation strategies for these key species. Furthermore, different research areas such as ecology and taxonomy could benefit themselves by the identification of these plants. DNA barcoding methods would be especially useful to identify samples lacking flowers, otherwise virtually impossible to identify without molecular tools.

References

Alves TLS, Chauveau OC, Eggers L, Souza-Chies (in press) Species discrimination in *Sisyrinchium* (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus. *Molecular Ecology Resources*.

Ashfaq M, Asif M, Anjum ZI, Zafar Y.(2013) Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Molecular Ecology Resources*, **13**, 573–582.

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.

Chauveau O, Eggers L, Souza-Chies TT, Nadot S (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of Botany*, **110**, 713–729.

Deble LP (2010) *Herbertia Zebrina* (Iridaceae, Tigridieae, Cipurinae) a new species from Rio Grande do Sul State (Brazil). *Darwiniana*, **48**, 93–96.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One*, **3**, e2802.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.

Goldblatt P (1990) Phylogeny and classification of Iridaceae. *Annals of the Missouri Botanical Garden*, **77**, 607-627.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C, Chave J (2009) Identification of Amazonian Trees with DNA Barcodes. *PLoS One*, **4**, e7483.

Hebert PDN, Cywinska A, Ball SL, DeWaard J (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan RS, Chase MW, Gaudeul M, Hollingsworth PM (2009) Selecting barcoding *loci* for plants: evaluation of seven candidate *loci* with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, **9**, 439–457.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One*, **6**, e19254.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary sudies. *Molecular Biology and Evolution*, **23**, 254–267.

Iganci JRV, Heiden G, Miotto STS, Pennington T (2011) Campos de Cima da Serra: the Brazilian Subtropical Highland Grasslands show an unexpected level of plant endemism. *Botanical Journal of the Linnean Society*, **167**, 378–393.

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjurb O, Berminghamb E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*, **106**, 18621–18626.

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS One*, **5**, e15409.

Lahaye R , van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA*, **105**, 2923–2928.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW (2011) Comparative analysis of a large dataset indicates that internal transcribed

spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, **108**, 19641–19646.

Li HQ, Chen JY, Wang S, Xiong SZ (2012) Evaluation of six candidate DNA barcoding *loci* in *Ficus* (Moraceae) of China. *Molecular Ecology Resources*, **12**, 783–790.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Maia VH, Mata CS, Franco LO, Cardoso MA, Cardoso SR, Hemerly AS, Ferreira PC (2012) DNA barcoding Bromeliaceae: achievements and pitfalls. *PLoS One*, **7**, e29877.

Meier R, Kwong S, Vaidya G, Nq PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Newmaster SG, Fazekas AJ, Steeves RA, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*, **8**, 480–490.

Nieto Feliner G, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution*, **44**, 911-919.

Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian *Paphiopedilum* species. *Molecular Ecology Resources*, **12**, 82–90.

Ren BQ, Xiang XG, Chen ZD (2010) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. *Molecular Ecology Resources*, **10**, 594–605.

Rodriguez A & Sytsma KJ (2006) Phylogeny of the ‘"tiger-flower" group (Tigirideae: Iridaceae): molecular and morphological evidence. *Aliso*, **22**, 412–424.

Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Datt B, Bag SK, Singh PK, Nair NK, Husain T, Tuli R (2010) Universal plant DNA barcode *loci* may not work in complex groups: A case study with Indian *Berberis* species. *PLoS One*, **5**, e13674

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120–1136.

Seberg O, Petersen G (2009) How many *loci* does it take to DNA barcode a *Crocus*? *PLoS One*, **4**, e4598.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Shilling EE, and R. L. Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.

Spooner DM (2009) DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *American Journal of Botany*, **96**, 1177–1189.

Starr JR, Naczi RF, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**, 151–163.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.

Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE (2012) DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources*, **12**, 620–633.

Wang W, Wu Y, Yan Y, Ermakova M, Kerstetter R, Messing, J (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots. *BMC Plant Biology*, **10**, 205.

Whitlock BA, Hale AM, Groff PA (2010) Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS One*, **5**, e11533.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. (2009) The frequency of polyploid speciation in vascular plants.*Proceedings of the National Academy of Sciences, USA*, **106**, 13875–13879.

Yang JB, Wang YP, Möller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). *Molecular Ecology Resources*, **12**, 267–275.

Yu J, Xue JH, Zhou, SL (2011a) New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution*, **49**, 76–181.

Yu Y, Than C, Degnan JH, Nakhleh L (2011b) Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic biology*, **60**, 138–149.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Molecular Ecology Resources*, **12**, 98–108.

Supplementary material

Table S1. Information about the primers used for amplification and sequencing in this study.

Barcoding region	Annealing temperature	Primer name	Primer function	Direction	Sequence (5'-3')	Reference
rbcL	56 ºC	rbcLa_f	PCR and sequencing	Forward	ATGTCACCACAAACAGAGACTAAAGC	(Levin, 2003)
	56 ºC	rbcLa_rev	PCR and sequencing	Reverse	GTAAAATCAAGTCCACCRCG	(Kress & Erickson, 2007)
matK	53 ºC	matK-Xf	PCR and sequencing	Forward	TAATTTACGATCAATTCATTC	http://www.kew.org/ barcoding/protocols.html
	53 ºC	matK-r2	Sequencing	Reverse	AGTTTGATAATTGGTTTATATG	(Chauveau *et al.*, 2012)
	53 ºC	matK-f2	Sequencing	Forward	CATATAAACCAATTATCAAAC	(Chauveau *et al.*, 2012)
	53 ºC	matK-3.2r	PCR and sequencing	Reverse	CTTCCTCTGTAAAGAATTC	http://www.kew.org /barcoding/protocols.html
trnH-psbA	56 ºC	psbA	PCR and sequencing	Forward	GTTATGCATGAACGTAATGCTC	(Shaw *et al.*, 2005)
	56 ºC	trnH^{GUG}	PCR and sequencing	Reverse	CGCGCATGGTGGATTCACAATCC	(Shaw *et al.*, 2005)
ITS	55 ºC	ITS tig	PCR and sequencing	Forward	GATCGGACGAACGCGAAC	This study
	55 ºC	ITS 4r_1	PCR and sequencing	Reverse	GTAAAAATCAAGTCCACCRCG	This study

Supplementary material references

Chauveau O, Eggers L, Souza-Chies TT, Nadot S (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of Botany*, **110**, 713–729.
http://www.kew.org/barcoding/protocols.html
Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *Plos One*, **2**, e508.
Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KJ (2003) Family-level relationships of Onagraceae based on chloroplast *rbcL* and *ndhF* data. *American Journal of Botany*, **90**, 107-115.
Shaw J, Lickey EB, Beck JT, Susan B. Farmer, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

Conclusão

Considerações finais

Nesta tese, os principais marcadores moleculares utilizados para a aplicação de *DNA barcoding* em espécies vegetais foram testados no intuito de identificar espécies pertencentes às tribos Sisyrinchieae e Tigridieae (Iridaceae: Iridoideae), dois grupos de plantas reconhecidos por apresentarem significativa complexidade taxonômica (Chauveau *et al*., 2011, 2012; Karst & Wilson, 2011; Rodriguez & Sytsma, 2006). Nossos resultados indicaram que ambos os grupos são difíceis de caracterizar usando as regiões genômicas recomendadas pelo *Consortium for the Barcoding of Life* (CBOL) (CBOL, 2009). A variabilidade inerente a cada marcador utilizado e as taxas de identificação das espécies estudadas utilizando tais marcadores foram consideradas relativamente baixas. Ainda assim, alguns *loci* foram úteis na identificação e caracterização de algumas linhagens, principalmente os espaçadores nrITS em Sisyrinchieae e o gene *matK* em Tigridieae.

Os resultados demonstraram que as diferentes regiões avaliadas respondem por taxas de reconhecimento de espécies de forma distinta conforme o grupo taxonômico em questão. Em Sisyrinchieae, amplamente amostrada e representada majoritariamente pelo gênero *Sisyrinchium* L., os estudos foram concentrados em dois marcadores plastidiais (*matK* e *trnH-psbA*) juntamente com o nrITS. Este último apresentou nitidamente os melhores índices de identificação de espécies. Entretanto, um máximo de cerca de 50% das espécies puderam ser identificadas através da combinação de marcadores, um valor considerado modesto. Os marcadores plastidiais *matK* e *trnH-psbA* mostraram-se pouco polimórficos e portanto uma capacidade de reconhecimento de espécies muito pequena. O melhor desempenho quando um único *locus* foi analisado foi alcançado pelo nrITS, sendo este substancialmente superior aos marcadores plastidiais. Assim, esta região foi

considerada a mais adequada para o *DNA barcoding* de membros da tribo Sisyrinchieae até o momento.

Uma amostragem menor, mais abrangente em número de gêneros e geograficamente concentrada foi utilizada na aplicação de *DNA barcoding* para a tribo Tigridieae. Os resultados quanto ao poder de identificação atingiram magnitudes similares, sendo que novamente cerca de 50% dos táxons foram corretamente reconhecidos, mas desta vez sem a utilização de marcadores combinados. Entre os quatro marcadores testados (*rbcL*, *matK*, *trnH-psbA* e nrITS), o de melhor desempenho foi o gene *matK*, provavelmente porque o *locus* nrITS, normalmente reconhecido como o mais polimórfico (Hollingsworth *et al*., 2011; Li *et al*., 2011), apresentou universalidade limitada em Tigridieae. Os *loci rbcL* e *trnH-psbA* mostraram-se pouco polimórficos em Tigridieae, com taxas de identificação insuficientes, sendo assim, o gene *matK* foi estabelecido como aregião mais adequada para o reconhecimento de espécies desta tribo até agora.

Os resultados obtidos nesta tese podem ser comparados a outros estudos realizados em Iridaceae e em outras linhagens de monocotiledôneas. No único trabalho de *DNA barcoding* publicado até o momento para Iridaceae, Seberg & Petersen (2009), usando marcadores exclusivamente plastidiais, obtiveram taxas de identificação um pouco superiores às verificadas para Sisyrinchieae e Tigridieae ao analisarem o gênero *Crocus* L. (Iridaceae: Crocoideae). Entretanto, esse trabalho utilizou uma amostragem intraespecífica limitada e combinação sequencial de vários *loci*, o que impede maiores comparações com nossos resultados. Ainda assim, as taxas de identificação obtidas sem combinar *loci* em *Crocus* (Seberg & Petersen, 2009) foram apenas ligeiramente superiores às obtidas para Sisyrinchieae e Tigridieae. Na família Orchidaceae, família filogeneticamente próxima a Iridaceae na ordem Asparagales, a combinação das regiões ITS e *matK* é considerada s

mais efetiva para identificar espécies de *Olcoglossum* Schltr (Xiang *et al*. 2011). Porém, foi ressaltada a necessidade de maior poder de resolução pelo uso de mais marcadores para uma parcela das espécies estudadas que não puderam ser efetivamente identificadas em razão de sua radiação recente (Xiang *et al*. 2011). Em contrapartida, sucesso pleno de identificação foi alcançado para o gênero *Paphiopedilum* Pfitzer com o uso exclusivo do *locus matK* (Parveen *et al*., 2012), caracterizando um caso incomum de alto desempenho de identificação em Asparagales. Na mais distantemente relacionada ordem Poales, estudos de *DNA barcoding* em Bromeliaceae (Maia *et al*., 2012) e no gênero *Carex* L. (Cyperaceae) (Starr *et al*., 2009) apresentaram taxas de identificação de espécies insuficientes Esse conjunto de resultados centrados em monocotiledôneas indica que os trabalhos de *DNA barcoding* nesse grupo taxonômico apresentam, em sua maioria, limites de identificação. Na verdade, o sucesso na identificação de espécies vegetais costuma variar fortemente de modo geral, sendo influenciado pelo grupo taxonômico avaliado ou ainda com a área geográfica em estudo.

A história evolutiva dos grupos em análise parece ter impacto significativo nos resultados obtidos. Existem evidências de que uma parcela substancial dos táxons de Tigridieae e Sisyrinchieae diversificou rapidamente, muitas vezes por influência de alterações na morfologia floral em resultado a adaptações aos polinizadores (Chauveau *et al*., 2011, 2012; Rodriguez & Sytsma, 2006). Além disso, a ocorrência de hibridações naturais em *Sisyrinchium* (Cholewa & Henderson, 1984; Henderson, 1976; Yamaguchi & Hirai, 1987) e poliploides em Tigridieae (Souza-chies *et al*., 2012) e também em algumas linhagens de *Sisyrinchium* (Souza-Chies *et al*., 2012; Tacuatiá *et al*., 2012a,b) são considerados processos que costumam comprometer a eficácia dos métodos em *DNA barcoding* (Fazekas *et al*., 2009). É interessante notar que boa parte das espécies de

Sisyrinchium e gêneros próximos são indistinguíveis na ausência de flores, sobretudo dentro de cada um dos clados estabelecidos por Chaueveau *et al*. (2011), onde a semelhança pode ocorrer inclusive em caracteres florais. De modo convergente, boa parte das espécies de Tigridieae assemelha-se morfologicamente em estado vegetativo (Rodriguez & Sytsma, 2006). Os membros de Tigridieae são, em geral, ervas bastante pequenas, muitas vezes constituindo plantas reduzidas a um bulbo diminuto e algumas poucas folhas. No entanto, quando inicia o florescimento, essas plantas impressionam pela exuberante variedade de flores de diferentes cores, tamanhos e formas. Há evidências de que a morfologia floral constitui um importante fator envolvido na velocidade de diversificação neste grupo (Chauveau *et al*., 2012). As filogenias publicadas até o momento indicam que tanto Sisyirinchieae quanto as espécies de Tigridieae analisadas são táxons filogeneticamente próximos (Chauveau *et al*., 2012). Assim, é de se esperar que a proximidade morfológica dentro de cada um dos dois grupos em análise esteja associada em algum grau à homogeneidade encontrada para a maioria dos marcadores usados para *DNA barcoding* nesta tese.

Este trabalho forneceu o primeiro conjunto significativo de dados de *DNA barcoding* aplicados a dois importantes grupos de Iridaceae distribuídos predominantemente no continente americano, e tais dados permitiram a avaliação sobre a capacidade e as limitações das técnicas de identificação de espécies com as ferramentas de *DNA barcoding*. Os táxons em estudo são considerados de difícil identificação, justificando ainda mais o uso de métodos de molecularespara o reconhecimento de espécies. Estudos adicionais buscando maior número de marcadores moleculares, amostragem intrapopulacional mais densa e o uso de novos métodos de identificação merecem consideração como alternativas para melhorar os níveis de resolução do *DNA barcoding*

nestes grupos de Iridaceae. A despeito de algumas limitações inerentes do *DNA barcoding* em Sisyrinchieae e Tigridieae, este trabalho certamente contribui para situações que necessitem a identificação destas espécies, inclusive em seu estado vegetativo. Assim, diferentes áreas como a ecologia, a taxonomia e a biologia da conservação podem se beneficiar do uso destes métodos incorporando-os como uma nova ferramenta em procedimentos que necessitem de identificação taxonômica.

2. Referências

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.

Chauveau O, Eggers L, Raquin C, Silvério A, Brown S, Couloux A, Cruaud C, Kaltchuk-Santos E, Yockteng R, Souza-Chies TT, Nadot S (2011) Evolution of oil producing trichomes in *Sisyrinchium* (Iridaceae): insights from the first comprehensive phylogenetic analysis of the genus. *Annals of Botany*, **107**, 1287–1312.

Chauveau O, Eggers L, Souza-Chies TT, Nadot S (2012) Oil-producing flowers within the Iridoideae (Iridaceae): evolutionary trends in the flowers of the New World genera. *Annals of Botany*, **110**, 713–729.

Cholewa AF, Henderson DM (1984) Biosystematics of *Sisyrinchium* section *Bermudiana* (Iridaceae) of the Rocky Mountains. *Brittonia*, **36**, 342–363.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.

Henderson DM (1976) A biosystematic study of Pacific Northwestern blue-eyed grasses (*Sisyrinchium*, Iridaceae). *Brittonia*, **28**, 149–176.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One*, **6**, e19254.

Karst L, Wilson CA (2012) Phylogeny of the New World genus *Sisyrinchium* (Iridaceae) based on analyses of plastid and nuclear DNA sequence data. *Systematic Botany*, **37**, 87–95.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, **108**, 19641–19646.

Maia VH, Mata CS, Franco LO, Cardoso MA, Cardoso SR, Hemerly AS, Ferreira PC (2012) DNA barcoding Bromeliaceae: achievements and pitfalls. *PLoS One*, **7**, e29877.

Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian *Paphiopedilum* species. *Molecular Ecology Resources*, **12**, 82–90.

Rodriguez A & Sytsma KJ (2006) Phylogeny of the ‘"tiger-flower" group (Tigirideae: Iridaceae): molecular and morphological evidence. *Aliso*, **22**, 412–424.

Seberg O, Petersen G (2009) How many *loci* does it take to DNA barcode a *Crocus*? PLoS One, **4**, e4598.

Souza-Chies TT, Santos EK, Eggers L, Flores AM, Stiehl-Alves EM, Fachinetto J, Lustosa J, Corrêa LB, Tacuatiá LO, Piccoli P, Miz RB (2012) Studies on diversity and evolution of Iridaceae species in southern Brazil. *Genetics and Molecular Biology*, **35**, 1027–1035.

Starr JR, Naczi RF, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**, 151–163.

Tacuatiá LO, Eggers L, Kaltchuk-Santos E, Souza-Chies TT (2012a) Population genetic structure of *Sisyrinchium micranthum* Cav. (Iridaceae) in Itapuã State Park, Southern Brazil. *Genetics and Molecular Biology*, **35**, 99–105.

Tacuatiá LO, Souza-Chies TT, Flores AM, Eggers L, Siljak-Yakovlev S, Kaltchuk-Santos E. (2012b) Cytogenetic and molecular characterization of morphologically variable *Sisyrinchium micranthum* (Iridaceae) in southern Brazil. *Botanical Journal of the Linnean Society*, **169**, 350–364.

Xiang XG, Hu H, Wang W, Jin XH (2011) DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. *Molecular Ecology Resources*, **11**, 1012–1021.

Yamaguchi H, Hirai S (1987) Natural hybridization and flower color inheritance in *Sisyrinchium rosulatum* Bickn.. *Weed Research*, **32**, 38–45.