RESEARCH ARTICLE

The use of the hypervariable P8 region of trnL (UAA) intron for identification of orchid species: Evidence from restriction site polymorphism analysis

Rajkumar Kishor*, G. J. Sharma

Department of Life Sciences, Manipur University, Imphal, Manipur, India

* Current address: Kwaklei and Khonggunmelei Orchids Pvt. Ltd., Sagolband Vijaygovind, Imphal, Manipur, India

Abstract

The P8 stem-loop region of the trnL intron, which is known to be hypervariable in size with multiple repeat motifs and created difficulties in alignment, is always excluded in phylogenetic as well as barcode analyses. This region was investigated for species discrimination in 98 taxa of orchids belonging to the tribe Vandeae using in silico mapping of restriction site polymorphism. The length of the P8 regions varied from 200 nucleotides in Aerides rosea to 669 nucleotides in Dendrophylax sallei. Forty two taxa had unique lengths, while as many as eight shared a common length of 521 nucleotides. Of the 35 restriction endonucleases producing digestions in the P8 regions, three, viz., AgsI, ApoI and TspDTI turned out to have recognition sites across all the 98 taxa being studied. When their restriction data were combined, 92 taxa could be discriminated leaving three taxon pairs. However, Acampe papillosa and Aeranthes arachnites despite having similar restriction sites differed in their P8 lengths. This is the first report on thorough investigation of the P8 region of trnL intron for search of species specific restriction sites and hence its use as a potential plant DNA barcode.

Introduction

For the past few decades there has been a hunt for a short DNA segment which can be used as a universal marker, popularly termed as DNA Barcode, for identification of faunal and floral species inhabiting this planet. For animal identification the mitochondrial gene cytochrome c oxidase subunit 1 (COX1) has proved successful [1] however, there is still difficulty in fixing a universal barcode for plants because of a more complex genetic background. Various nuclear and plastid coding and non-coding loci, viz., ITS1, ITS2, accD, matK, ndhJ, rpoB, rpoC1, ycf5, atpF-H, psbK-I, rbcL, rbcLa, trnH-psbA, trnL-F, etc. have been tested for validation either singly or in combination of two or more loci [2–4].

The chloroplast trnL-F non-coding region includes the trnL(UAA) intron ranging from 350 to 600 base pairs (bp) and the intergeneric spacer between trnL(UAA) 3′ exon and the
The *trn*L(UAA) intron interrupts the anticodon loop of the tRNA$_{Leu}$, which is encoded in the large single copy region of the plastid genome. In the chloroplast DNA, *trn*L is the only Group I intron region having conserved secondary structure [7,8] with alternation of conserved and variable regions [9]. They are capable of catalyzing their own splicing from the flanking exons. The secondary structure of *trn*L intron contains regions of complimentary sequences that form nine stem-loop structures (P1-P9) [10]. Within these stems there are four regions (P, Q, R and S) conserved in primary sequences among all group I introns [11] and they are known as the catalytic core [12]. The P8 stem-loop region of the *trn*L intron is known to be hypervariable in size with multiple repeat motifs [12,13] and created difficulties in alignment. Therefore, this region is always excluded in phylogenetic as well as barcode analyses [14–17].

Orchidaceae comprises of 850 genera and 20,000 species which are arranged in five subfamilies, 22 tribes and 70 sub-tribes [18]. The tribe Vandeeae consists of five sub-tribes, 139 genera with 2600 species of monopodial epiphytes distributed in tropical America, tropical and southern Africa, tropical and sub-tropical Asia, eastern Australia and Tasmania, much of the tropical Pacific south to New Zealand and east to Tahiti [19]. Considering the hypervariability to be an intrinsic property of the P8 region of *trn*L(UAA) intron, the present investigation was undertaken to find out whether the locus had any potential in discrimination of closely related species belonging to the tribe Vandeeae based on in silico restriction site polymorphism analysis.

Materials and methods

Gene sequences

One hundred and twenty five sequences of tRNA$_{Leu}$ (*trn*L) gene for orchids belonging to tribe Vandeeae were downloaded from GenBank. These included ten sequences (GU185926, GU185928, GU185931, GU185932, GU185933, GU185934, GU185935, GU185938, GU185939 and GU185940) generated and submitted by the first author. Methods followed for DNA extraction and PCR amplification were given by Kishor and Sunitibala [20]. Sequencing was done using 3730 DNA Analyzer (Applied Biosystems, Warrington, UK) available at DNA Sequencing Facility, University of Delhi, South Campus. The primer pair ‘c’ (3’CGAAATCGGTAGACGCTACG5’) and ‘d’ (3’GGGGATAGAGGGACTTTGAAC5’) [5] were used for amplification as well as sequencing.

In silico analysis of restriction site polymorphism for species discrimination

The *trn*L intron borders were delimited by identifying the binding sites of the primers ‘c’ and ‘d’. We looked for the P8 region by delimiting the borders following the method of Borsch et al. [13] and by considering the secondary structures of *trn*L of *Campylopus flexuosus* [21] and *Nymphaea odorata* [13]. The correctness of the borders was again verified by considering the secondary structures of the P8 regions drawn for *Aerides odorata*, *A. sukauensis* and *A. kraibensis* [16].

In silico restriction mapping was done using the online software *RestrictionMapper* Version 3 (http://restrictionmapper.org/). Selection was made to include all commercially available restriction enzymes and find the base pair position where each enzyme cut the *trn*L P8 region for all the taxa. The program was set to allow providing maximum cuts with a minimum site length of 5 nucleotides. As many as 35 restriction endonucleases were found to have recognition sites in the P8 regions; however, only four, viz., *Agl*, *Apol*, TspDTI and VspI (five and six
base cutters) were selected for the analysis because each of them had their recognition sites in the region across all the taxa.

The in silico mapped restrictions produced by the four enzymes were scored by presence (1) or absence (0). Further analyses were done using the computer package NTSYS [22]. Similarity matrix was generated using the program Qualitative in NTSYS 2.20e package. This matrix was subjected to the unweighted pair group method with arithmetical averages (UPGMA). Cluster analysis was performed on the similarity matrix with the SAHN program using UPGMA and the dendrograms were generated with the TREE program. Analysis was done for each enzyme as well as for combination of two, three and four of them.

Results

Length variation of trnL P8 region

The downloaded trnL sequences for the 125 accessions contained many with incomplete sequences. After establishing the borders, only 98 taxa with complete sequence of the P8 region could be selected for the analysis. The length of the P8 regions varied from 200 nucleotides (Aerides rosea) to 669 nucleotides (Dendrophyllax sallei) (Fig 1, Table 1). Forty two taxa had unique lengths, while as many as eight shared a common length of 521 nucleotides (Aeranthes arachnites, Aeranthes grandiflora, Angraecum conchiferum, Angraecum dives, Angraecum leonis, Bonnieria appendiculata, Phalaenopsis amboinensis, and Phalaenopsis modesta). Within the genus Phalaenopsis, which is represented by 23 species in the present study, length variation of the P8 regions ranged from 494 (P. tetraspis) to 538 nucleotides (P. lindenii); while in eight taxa of Vanda, it ranged from 471 (V. testaceae) to 562 nucleotides (V. griffithii).

In silico analysis of restriction site polymorphism

The result of the in silico restriction site analysis is presented in Table 1. Of all the commercially available restriction endonucleases being tried for the in silico restriction mapping using the online software 'RestrictionMapper', 35 endonucleases (AflI, Accl, AgsI, Apol, AsuII,
Table 1. Sequence lengths of \textit{trnL} P8 regions of 98 orchid taxa along with the restriction sites for the endonucleases \textit{AgsI}, \textit{ApoI}, \textit{TspDTI} and \textit{VspI}.

| Taxa                | GenBank Accession No. | Length of P8 region (nucleotides) | Restriction endonucleases and their respective restriction sites (nucleotides) |
|---------------------|-----------------------|-----------------------------------|--------------------------------------------------------------------------------|
| \textit{Acampe ochracea} | DQ091438              | 421                               | 259, 265, 293, 357, 400, 406                                                   |
| \textit{Acampe papillosa} | DQ091439              | 420                               | 259, 265, 293, 357, 400                                                       |
| \textit{Aeranthes arachnites} | DQ091536              | 521                               | 259, 265, 293, 357, 400                                                       |
| \textit{Aeranthes grandiflora} | DQ091537              | 521                               | 238, 400, 428, 434, 452, 500                                                  |
| \textit{Aerides crassifolia} | EF670402              | 492                               | 195, 397, 403, 428, 471                                                       |
| \textit{Aerides crispa} | EF670407              | 484                               | 189, 361, 389, 395, 412, 420, 463                                             |
| \textit{Aerides krabiensis} | EF670404              | 217                               | 196                                                                           |
| \textit{Aerides multiflora} | DQ194983              | 492                               | 195                                                                           |
| \textit{Aerides odorata} | EF670389              | 450                               | 166, 327, 355, 361, 386, 429                                                  |
| \textit{Aerides rosea} | EF670405              | 200                               | 138, 179                                                                      |
| \textit{Aerides sakaensis} | EF670408              | 479                               | 195, 356, 384, 390, 415, 458                                                   |
| \textit{Ancistrorhynchus cephalotes} | EF670435              | 619                               | 345, 526, 532, 598                                                           |
| \textit{Angraecum calceolus} | DQ091546              | 505                               | 238, 420, 484                                                                 |
| \textit{Angraecum conchiferum} | DQ091537              | 521                               | 238, 428, 434, 500                                                           |
| \textit{Angraecum divae} | DQ091547              | 521                               | 238, 400, 428, 434, 500                                                       |
| \textit{Angraecum eichleri} | AF506341              | 639                               | 145, 238, 446                                                                 |
| \textit{Angraecum florulentum} | DQ091540              | 513                               | 237, 392, 420, 426, 492                                                       |
| \textit{Angraecum leonis} | DQ091551              | 521                               | 238, 400, 428, 434, 500                                                       |
| \textit{Angraecum magdalenae} | AF519973              | 616                               | 334, 490, 524, 595                                                           |
| \textit{Angraecum rutenbergianum} | DQ091547              | 521                               | 238, 400, 428, 434, 500                                                       |
| \textit{Arachnis labrosa} | GU185926              | 478                               | 195, 357, 385, 391, 416, 459                                                   |
| \textit{Ascochilus curvifolium} | DQ091550              | 513                               | 237, 392, 420, 426, 492                                                       |
| \textit{Aserola leonis} | DQ091551              | 521                               | 238, 400, 428, 434, 500                                                       |
| \textit{Bonniera appendiculata} | DQ091154              | 521                               | 238, 428, 434, 451, 500                                                       |
| \textit{Christensonia vietnamica} | EF670413              | 497                               | 200, 367, 395, 401, 426, 476, 482                                             |
| \textit{Cleisostoma racemiferum} | GU185928              | 525                               | 231, 402, 430, 436, 461, 504, 510                                              |
| \textit{Cryptopus elatus} | DQ091450              | 523                               | 240, 402, 430, 436, 453, 502                                                   |
| \textit{Cryptopus paniculatus} | DQ091451              | 514                               | 233, 393, 421, 427, 444, 493                                                  |
| \textit{Dendrophyllus sallei} | AY147234              | 669                               | 186, 379, 590, 582, 648                                                       |
| \textit{Dimorphorchis rossii var. graciliscapace} | EF670429              | 519                               | 214, 396, 424, 430, 455, 498, 504                                             |
| \textit{Holcochilus kimbballianum} | EF670419              | 629                               | 218, 506, 534, 540, 565, 608                                                   |

(Continued)
### Table 1. (Continued)

| Taxa                          | GenBank Accession No. | Length of P8 region (nucleotides) | Restriction endonucleases and their respective restriction sites (nucleotides) | Accession No. | Length of P8 region (nucleotides) | Restriction endonucleases and their respective restriction sites (nucleotides) |
|-------------------------------|-----------------------|-----------------------------------|--------------------------------------------------------------------------------|---------------|-----------------------------------|--------------------------------------------------------------------------------|
| Holcoglossum subulatifolium   | EF670409              | 467                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Jumellea maxillarioides       | DQ091554              | 506                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Jumellea sagittata            | DQ091555              | 508                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Lemurochis madagascariensis   | DQ091556              | 528                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Listrostachys pertusa         | DQ091509              | 524                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Luisia trichorrhiza           | EF670428              | 605                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Luisia tristis               | EF670426              | 596                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Microcoelia stolzii           | DQ091350              | 480                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Neofinetia falcata           | EF670421              | 497                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Oeonia rosea                 | DQ091452              | 516                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Paraphalaenopsis labukenensis | EF670425              | 507                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Pelatantheria insectifera    | GU185931              | 482                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis ambomensis      | AY273643              | 521                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis bellina         | AY273632              | 495                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis celebensis      | AY265799              | 507                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis corningiana     | AY273670              | 508                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis cornu-cervi     | AY265751              | 501                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis deliciosa       | DQ091444              | 415                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis inscriptinosensis| AY273673              | 498                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis lamelligera     | AY273679              | 501                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis lindenii        | AY273649              | 538                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis lucdemaniana    | AY273640              | 517                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis maculata        | AY273641              | 519                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis manii           | AF519969              | 523                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis mariae          | AY265770              | 519                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis micholitzi      | AY273701              | 505                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis modesta         | AY265793              | 521                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Phalaenopsis pantherina      | AY273666              | 501                               | Aci (TTSSA) 146, 181, 188, 390, 426, 447 27, 193, 213, 293, 349, 407 121, 224 |
| Taxa                        | GenBank Accession No. | Length of P8 region (nucleotides) | Restriction endonucleases and their respective restriction sites (nucleotides) |
|----------------------------|-----------------------|------------------------------------|--------------------------------------------------------------------------------|
| *Phalaenopsis philippinensis* | AY273656              | 511                                | Acl (TTSSA) 192, 388, 416, 422, 447, 490, 496                                |
| *Phalaenopsis pulchra*      | AY273639              | 517                                | Apol (RAATTY) 151, 186, 193, 434, 470                                      |
| *Phalaenopsis schilleriana* | AY265781              | 511                                | TipDTI (ATGAA) 27, 198, 218, 393, 451                                       |
| *Phalaenopsis stuartiana*   | AY273654              | 507                                | Vsp (ATTAAT) 126, 245, 259                                                  |
| *Phalaenopsis sumatrana*    | AY273695              | 508                                |                                                                ............... |
| *Phalaenopsis tetraspis*    | AY265784              | 494                                |                                                                ............... |
| *Phalaenopsis violacea*     | AY273635              | 507                                |                                                                ............... |
| *Pomatocalpa diffusum*      | EF670432              | 471                                |                                                                ............... |
| *Pomatocalpa spicatum*      | EF670431              | 490                                |                                                                ............... |
| *Rangaeris amaniensis*      | DQ091512              | 517                                |                                                                ............... |
| *Rangaeris rhipsalisocia*   | DQ091511              | 515                                |                                                                ............... |
| *Renanthera imischootiana*  | GU185932              | 472                                |                                                                ............... |
| *Renanthera mutatina*       | AY273688              | 498                                |                                                                ............... |
| *Renanthera histrionica*    | AY273692              | 496                                |                                                                ............... |
| *Rhipidoglossum kamerunense*| DQ091492              | 512                                |                                                                ............... |
| *Rhipidoglossum rutulum*    | DQ091494              | 522                                |                                                                ............... |
| *Rhipidoglossum subsimplex* | DQ091496              | 512                                |                                                                ............... |
| *Rhynchostylis gigantea*    | EF670411              | 479                                |                                                                ............... |
| *Rhynchostylis retusa*      | EF670424              | 501                                |                                                                ............... |
| *Sedirea japonica*          | EF670433              | 473                                |                                                                ............... |
| *Seidenfadenia mitrata*     | EF670414              | 490                                |                                                                ............... |
| *Smitinandia micrantha*     | EF670415              | 481                                |                                                                ............... |
| *Sobernikoffia hambertiana* | DQ091558              | 514                                |                                                                ............... |
| *Trichoglossis atropurpurea*| DQ091440              | 430                                |                                                                ............... |
| *Trichoglossis bipunctata*  | EF670430              | 531                                |                                                                ............... |
| *Tridactyle crassifolia*    | DQ091515              | 530                                |                                                                ............... |
| *Tridactyle filifolia*      | DQ091516              | 530                                |                                                                ............... |
| *Tridactyle tanneri*        | DQ091520              | 518                                |                                                                ............... |
| *Vanda alpina*              | GU185933              | 526                                |                                                                ............... |
| *Vanda coerulea*            | GU185935              | 549                                |                                                                ............... |

(Continued)
Bci VI, Bcl I, Bda I, Bgl II, BsaAI, BsaBI, BsmI, CapCI, Eco57I, Eco57MI, EcoRI, EcoRV, HaeIV, MfeI, MboII, MslI, NdeI, PsiI, SmI, SmaBI, SpeI, TfiI, TspDTI, TspGW I, TspRI, VspI, XbaI, XhoII and XmnI) were found to have recognition sites in the P8 region. Number of restriction endonucleases digesting the region varied from eight in Aerides krabiensis, A. odorata and A. rosea to 25 in Microcoelia stolzii (Fig 2). Again, from amongst the 35 restriction endonucleases, only three, viz., AgsI, Apol and TspDTI turned out to have recognition sites across all the 98 taxa being studied; while VspI had in all excepting Angraecum dives. Hence, the above four were selected for subsequent analysis of the restriction site polymorphism.

R = A or G, S = C or G, Y = C or T; while submitting to GenBank the trnL sequences Vanda motesiana* was entered as V. stangeana and V. stangeana* entered as V. tessellate

https://doi.org/10.1371/journal.pone.0196680.t001

[Vanderbilt University Libraries](https://doi.org/10.1371/journal.pone.0196680.g002)
Species discrimination by single restriction endonuclease

The restriction sites recognized by Agsl ranged from one (Aerides krabiensis and A. multiflora) to 14 (Phalaenopsis inscriptiosinensis) (Table 1). From the dendrogram constructed by the UPGMA method which showed similarity coefficient ranging from 0.91 to 1.00 (Fig 3), it was found that Agsl exhibited species specific restriction sites in 80 taxa; however, it could not provide discrimination for 18 taxa which included the taxon pairs Acampe papillosa and Aeranthes arachnites; Angraecum dives and Angraecum leonis; Phalaenopsis lamelligera and P. pantherina, P. lueddemanniana and P. pulchra; Phalaenopsis philippinensis and P. schilleriana; Phalaenopsis

Fig 3. UPGMA dendrogram based on SAHN clustering of data from in silico restriction site polymorphism analysis of the trnL(UAA) intron P8 region using the restriction endonuclease Agsl among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g003
maculata and P. mariae; Phalaenopsis amboinensis and P. modesta; Phalaenopsis corningiana and P. sumatrana; Pomatocalpa diffusum and Vanda testacea. Each of these taxon pairs shared the same Apol restriction sites.

The Apol recognized restriction sites ranged from two (in six taxa) to ten (Phalaenopsis inscriptiosinensis) (Table 1). As evidenced from the UPGMA dendrogram (Fig 4), of the 98 taxa analysed Apol digestion discriminated 77 species, however, it could not show species specific restriction sites in 21 taxa. The following taxon pairs or groups shared the same restriction sites, Acampe ochraceae, Acampe papillosa and Aeranthes arachnites, Angraecum calceolus and Jumellia maxillarioideae, Angraecum conchiferum, Angraecum dives, Angraecum leonis and Bonniera appendiculatula, Rhipidoglossum kamerunense and Rhipidoglossum subsimplex, Tridactyle crassifolia and Tridactyle filifolia, Phalaenopsis amboinensis and P. modesta, Phalaenopsis lamelligera and P. pantherina, Phalaenopsis corningiana and P. sumatrana, Phalaenopsis philippinensis and P. schilleriana. This dendrogram showed similarity coefficient ranging from 0.92 to 1.00.

TspDTI exhibited species specific restriction patterns for 78 taxa and could not discriminate 20 taxa (Table 1, Fig 5). These twenty included the taxon pairs or groups Acampe papillosa and Aeranthes arachnites; Christensonia vietnamica and Seidenfadenia mitrata; Phalaenopsis celebensis and Phalaenopsis stuartiana; Phalaenopsis philippinensis and P. schilleriana; Phalaenopsis corningiana and P. sumatrana; Vanda testacea, Vandopsis lissochiloides and Vandopsis gigantea; Rhipidoglossum kamerunense and Rhipidoglossum subsimplex; Tridactyle crassifolia and Tridactyle filifolia. The UPGMA dendrogram showed similarity coefficient ranging from 0.92 to 1.00.

Vsp1 also exhibited species specific restriction patterns for 66 taxa and could not discriminate 32 taxa (Table 1, Fig 6). These 32 include the taxon pairs or groups, Acampe ochraceae, Acampe papillosa and Aeranthes arachnites; Aerides multiflora, Phalaenopsis deliciosa and Tridactyle tanneri; Christensonia vietnamica and Seidenfadenia mitrata; Arachnis labrosa and Rynchostylysis gigantea; Vanda testacea, Vandopsis lissochiloides and Vandopsis gigantea; Angraecum conchiferum, Angraecum leonis and Bonniera appendiculatula; Listrostachys pertusa, Microcoelia stolzii, Rangaeris amanuensis, Rangaeris ripsalisoca, Tridactyle crassifolia and Tridactyle filifolia; Rhipidoglossum kamerunense and Rhipidoglossum subsimplex; Phalaenopsis philippinensis and P. schillerian; Phalaenopsis bellina and Phalaenopsis micholitzii; Phalaenopsis cornu-cervi and Phalaenopsis lamelligera; Phalaenopsis corningiana and P. sumatrana. The UPGMA dendrogram showed similarity coefficient ranging from 0.90 to 1.00.

Species discrimination using combined data of restriction endonucleases

Analysis of the in silico restriction mapping using the combined data of Agel and Apol resulted in discrimination of 86 taxa (Fig 7) as against their individual resolutions of 80 and 77 taxa respectively. Those which had exactly the same restriction sites included six taxon pairs, Acampe papillosa and Aeranthes arachnites; Angraecum dives and Angraecum leonis; Phalaenopsis lamelligera and P. pantherina; Phalaenopsis amboinensis and P. modesta; Phalaenopsis corningiana and P. sumatrana; Phalaenopsis philippinensis and P. schilleriana. The UPGMA dendrogram showed similarity coefficient ranging from 0.91 to 1.00.

Again, when Apol, Agel and TspDTI restriction data were combined, 92 taxa could be discriminated while the three taxon pairs which could not be discriminated from each other included Acampe papillosa and Aeranthes arachnites; Phalaenopsis corningiana and P. sumatrana; Phalaenopsis philippinensis and P. schilleriana (Table 1, Fig 8). The UPGMA dendrogram showed similarity coefficient ranging from 0.92 to 1.00.
Fig 4. UPGMA dendrogram based on SAHN clustering of data from in silico restriction site polymorphism analysis of the \textit{trnL(UAA)} intron P8 region using the restriction endonuclease Apol among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g004
Combination of data of all the four enzymes ApoI, Agsl, TspDTI and VspI also gave the same result as that of the above three enzymes with no discrimination of the following three pairs *Acampe papillosa* and *Aeranthes arachnites*; *Phalaenopsis corningiana* and *P. sumatrana*;
Fig 6. UPGMA dendrogram based on SAHN clustering of data from the in silico restriction site polymorphism analysis of the trnL(UAA) intron P8 region using the restriction endonuclease VspI among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g006
Fig 7. UPGMA dendrogram based on SAHN clustering of the combined data from in silico restriction site polymorphism analyses of the trnL(UAA) intron P8 region using the restriction endonuclease Agel and Apol among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g007
Fig 8. UPGMA dendrogram based on SAHN clustering of the combined data from in silico restriction site polymorphism analyses of the trnL(UAA) intron P8 region using the restriction endonuclease *AgeI*, *ApoI* and *TspDII* among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g008
Phalaenopsis philippinensis and P. schilleriana (Table 1, Fig 9). The UPGMA dendrogram showed similarity coefficient ranging from 0.92 to 1.00.

Discussion

The P8 stem-loop region is reported to be the most length-variable region of the trnL intron which is due to presence of repeats of various sizes [21]. These repeats together with the enormous length and sequence variation hampered alignment in most of the plant groups being studied for phylogeny or barcoding and therefore, this region is often excluded from analysis of trnL intron sequences [14–16]. Due to this exclusion, the results inferred from analysis of the trnl intron sequences do not represent the exact information it should naturally infer. By virtue of its hypervariability, the P8 region should be having certain information which may be analysed and interpreted for applying in species level identification or barcoding, if not for phylogenetic inference. As there is difficulty in sequence alignment of the P8 regions, the only other method for analysis should be the restriction analysis. In silico restriction site polymorphism was opted over the conventional PCR-Restriction Fragment Length Polymorphism (PCR-RFLP) as there were already a large number of trnl sequences deposited to GenBank. Moreover, the PCR-RFLP may not show the true picture in case two fragments of equal lengths coming from different parts of the sequence cannot be distinguished by observing the gel picture. For plant species identification PCR-RFLP is seldom used, however, a few investigators employed it for identification of mangrove and mangrove associate species [23], fine roots of trees from the Alps [24], Cinnamomum spp. [25], upland grassland species from roots [26] and Vasconcellea species [27], Dendrobium orchids [28].

Sequence length analysis

The tribe Vandeae (Family: Orchidaceae) is very robust consisting of 2600 species of monopodial epiphytes under 139 genera. Some previous investigators had already worked on phylogeny of these orchids using trnL intron sequences [15–17]. As an objective of the present investigation we aimed to discriminate closely related species belonging to a genus or a subtribe or a tribe. While searching the GenBank database, 125 accessions of trnL intron sequence for taxa, under the tribe Vandeae, was available when we started the investigation and they were retrieved for analysis. Upon further investigation 98 out of the 125 sequences had complete P8 hypervariable regions, and thus were considered for the final analysis. Sequence length variation of the P8 regions could discriminate 92 out of the 98 taxa of orchids being investigated. It was observed that the length of the P8 regions of all the taxa being varied from 200 (Aerides rosea) to 669 nucleotides (Dendrophylax sallei). Dendrophylax sallei had 30 more nucleotides longer than Luisia curtisii (Orchidaceae) which was reported earlier to be the longest recorded angiosperm P8 (639 nucleotides) [16]. As evidenced from our finding, those taxa sharing a common length did not necessarily belong to the same genus and that there could be great differences in P8 lengths within the same genus, which is in conformity with that of Kocyan et al. [16]. This difference in P8 lengths might be due to slipped-strand mispairing (SSM) resulting into high repetition of A motifs [21, 29].

Restriction site polymorphism analysis

Our study represents the first ever in silico restriction analysis of the P8 region of trnL in an attempt to utilize the genetic information present in it for a meaningful interpretation in species identification of a certain group of angiospermic plants. From all the UPGMA dendrograms generated in this investigation it is evidenced that species belonging to one genus are clustered with those belonging to other genus or genera, or in the other sense they are
Fig 9. UPGMA dendrogram based on SAHN clustering of the combined data from in silico restriction site polymorphism analyses of the trnL(UAA) intron P8 region using the restriction endonuclease AgII, Apol, TspDTI and VspI among the 98 taxa of orchids.

https://doi.org/10.1371/journal.pone.0196680.g009
haphazardly clustered owing to the hypervariable nature of the P8 regions. This showed that the present approach might not be applicable for phylogenetic inference of the taxa being investigated. However, since 95.9% of these taxa could be discriminated based on restriction site polymorphisms and sequence length data, this technique might be adopted for rapid species identification and hence as a plant DNA barcode.

So far there has not been much report on utilization of PCR-RFLP for identification of orchid species except for certain Thai Dendrobium orchids using rDNA-ITS and cpDNA regions [28]. Some investigators already showed the efficacy of double digestion or using two or more restriction endonucleases over single in generation of more polymorphic fragments in PCR-RFLP experiments. PCR-RFLP of the chloroplast trnS-psbC gene regions using a combination of two enzymes, HaeIII and MspI could successfully identify all the 119 accessions of millet into 7 species [30]. 579 grasses roots were distinguished to ten species using PCR-RFLP of trnL intron, with one or two enzyme digest [26]. Again, 16 taxa out of 30 tree species from the Alps were identified using PCR-RFLP with four restriction endonucleases TaqI, HinfI, RsaI and CfoI [24]. However, in our investigation, it was observed that as many as 35 restriction endonucleases had their recognition sites in the region. Combined data from analyses of the P8 regions with three restriction endonucleases Apol, Agsl and TspDTI, could discriminate 92 of the 98 taxa based on species specific restriction sites. Hence, the advantage of screening a large number of restriction endonucleases is required for higher success rate in species discrimination of the plant specimens being investigated.

Aerides rosea, having the shortest P8 length of 200 nucleotides, showed to have recognition sites of at least eight enzymes, while Dendrophylax salli despite having the longest P8 length (669 nucleotides) had recognition sites for only 20 enzymes. Microcoelia stolzii with a P8 length of 480 nucleotides had the maximum number (25) of restriction endonucleases cutting the region. Again, considering the number of recognition sites per restriction endonuclease for an individual taxon, Phalaenopsis inscriptiosinensis, with a P8 length of 498 nucleotides, had as many as 14 Agsl, 10 Apol, 10 TspDTI and 8 VspI recognition sites. Hence, our result showed that the longest P8 length neither had the maximum number of restriction endonuclease recognizing it nor maximum number of recognition sites for an individual enzyme.

It was observed that there were 20 restriction endonucleases having recognition sites in the P8 regions of both Acampe papillosa and Aeraanthes arachnites. These two taxa also exhibited the same number of restriction sites for all the twenty enzymes. The UPGMA trees drawn for each of the four enzymes as well as those for their combined data revealed them to have 100% genetic similarity. Hence, the only information to differentiate them as different species would be their P8 sequence lengths, 420 nucleotides for Acampe papillosa and 521 nucleotides for Aeraanthes arachnites.

Phalaenopsis corningiana and P. sumatrana are treated as two separate species [31–33]. Our investigation revealed that both of them had the same P8 length (508 nucleotides) and identical sequence and hence could not be distinguished from each other as they had similar restriction sites for all the 19 restriction enzymes. Tsai [34] also could not separate the two taxa based on information derived from nrITS, IGS and atpB-rbcL sequences. The only morphological differences between them are in callus and marking patterns on the petals. Distribution of P. sumatrana is widespread ranging from Myanmar, Thailand, Vietnam, Indonesia, Malaysia and the Philippines; whereas P. corningiana is restricted to Borneo and Sarawak. Phalaenopsis philippinensis and P. schilleriana are also endemic to Philippines and they have distinct distinguishing morphological characters. From our analysis, it is observed that both possessed identical P8 sequences and hence identical restriction endonucleases and their restriction sites. Whether these two taxon pairs should be treated as natural hybrids or ecotypes may need evidences from other coding, non-coding sequences or protein markers.
It was suggested that for a gene region to be practical as a DNA barcode the following three criteria must be fulfilled: (i) contain significant species-level genetic variability and divergence, (ii) possess conserved flanking sites for developing universal PCR primers for wide taxonomic application, and (iii) have a short sequence length so as to facilitate current capabilities of DNA extraction and amplification [3]. From our result, it is learnt that P8 regions of the 98 taxa contained good amount of genetic variation either in sequence length or restriction sites of the enzymes being used. Those taxa which could not be discriminated might require understanding of their maternal origin, in case of natural hybridization; plastid DNA barcodes will fail in case of natural hybrids. Second, the primer pair (c and d) used to amplify trnL intron is well conserved from bryophytes to angiosperms [35]. The present study also employed this primer pair for both PCR amplification as well as sequencing. Using these primers, the whole trnL region could be amplified and sequenced and from it an intact P8 region be retrieved easily, which is an advantage. Third, sequence lengths of the P8 regions observed in the present investigation ranged from 200 to 669 nucleotides which were short enough to be considered as DNA barcodes. The only disadvantage of the P8 region lie in their inability to be aligned due to hypervariability, and hence cannot be used for further processing using standardized phylogenetic or barcoding techniques.

Conclusions
A technique for molecular identification using sequence length variation and in silico restriction site polymorphism analysis of the trnL intron P8 sequence was developed and utilized to discriminate 94 out of 98 taxa of orchids to the level of species. Investigations using this technique for species level discrimination across all the angiospermic families may be tried. The four restriction endonucleases Apol, AgsI, TspDTI and VspI could be utilized for further analysis of the trnL intron P8 sequences of other uninvestigated orchid taxa either in silico or in PCR-RFLP. A plant DNA barcoding system using restriction site polymorphism of the trnL P8 region has not been suggested yet; and with this report there is high possibility of using this tool to establish a barcoding system.

Acknowledgments
This work was supported by the Council of Scientific and Industrial Research (CSIR), Government of India by granting Senior Research Associateship to Rajkumar Kishor (Grant No. SRA-8627A).

Author Contributions

Conceptualization: Rajkumar Kishor.
Data curation: Rajkumar Kishor.
Formal analysis: Rajkumar Kishor.
Funding acquisition: Rajkumar Kishor.
Investigation: Rajkumar Kishor.
Methodology: Rajkumar Kishor.
Supervision: G. J. Sharma.
Writing – original draft: Rajkumar Kishor.
Writing – review & editing: G. J. Sharma.
References

1. Hebert PDN, Ratnasingham S, De Waard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related animals. Proc Roy Soc London, Ser B, Biol Sci. 2003; 270: S96–S99.

2. Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinnan S, Petersen G, et al. A proposal for a standardized protocol to barcode all land plants. Taxon 2007; 56: 295–299.

3. Kress WJ, Erickson DL. DNA barcodes: Genes, genomics, and bioinformatics. Proc Natl Acad Sci U S A. 2008; 105: 2761–2762. https://doi.org/10.1073/pnas.0800476105 PMID: 18287050

4. Hollingsworth PM, Grahm SW, Little DP. Choosing and using a plant DNA Barcode. PLoS One. 2011; 6(5)e19254:1–13.

5. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol. 1991; 17: 1105–1109.

6. Gielly L, Taberlet P. The use of chloroplast DNA to resolve phylogenies: noncoding versus rbcL sequences. Mol Biol Evol. 1994; 11: 769–777. https://doi.org/10.1093/oxfordjournals.molbev.a040157 PMID: 7968490

7. Kress WJ, Erickson DL. DNA barcodes: Genes, genomics, and bioinformatics. Proc Natl Acad Sci U S A. 2008; 105: 2761–2762. https://doi.org/10.1073/pnas.0800476105 PMID: 18287050

8. Davies RW, Waring RB, Ray JA, Brown TA, Scazzochio C. Making ends meet—a model for RNA splicing in fungal mitochondria. Nature. 1982; 300: 719–724. PMID: 6757759

9. Michel F, Jacquier A, Dujon B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie. 1982; 64: 867–88. PMID: 6817818

10. Carlsward BS, Whitten WR, Williams NH, Bytebier B. Molecular phylogenetics of Vandeae (Orchidaceae) and the evolution of leaflessness. Am J Bot. 2006; 93:770–786. https://doi.org/10.3732/ajb.93.5.770 PMID: 18571438

11. Dressler RL. Phylology and classification of the orchid family. Dioscorides Press, Portland, OR. 1993.

12. Carlsward BS, Whitten WR, Williams NH, Bytebier B. Molecular phylogenetics of Vandeae (Orchidaceae) and the evolution of leaflessness. Am J Bot. 2006; 93:770–786. https://doi.org/10.3732/ajb.93.5.770 PMID: 18571438

13. Kocyan A, de Vogel EF, Conti E, Gravendeel B. Molecular phylogeny of Aerides (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. Mol Phylogenet Evol. 2008; 48: 422–444. https://doi.org/10.1016/j.ympev.2008.02.017 PMID: 18571438

14. Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH. Molecular phylogeny of Phalaenopsis Blume (Orchidaceae) on the basis of plastid and nuclear DNA. Plant Syst Evol. 2010; 288: 77–98.

15. Dressler RL. Phylology and classification of the orchid family. Dioscorides Press, Portland, OR. 1993.

16. Schuiteman A. Dendrobieae and Aeridinae. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, editors. Genera Orchidacearum: Volume 6: Epidendroidea, Part 3, Oxford University Press; 2014.

17. Kishor R, Sunitibala DH. Induction of multiple shoots in a monopodial orchid hybrid (Aerides vandarum Reichb.f x Vanda stangeana Reichb.f) using thidiazuron and analysis of their genetic stability. Plant Cell Tiss Organ Cult. 2009; 97: 121–129.

18. Schuiteman A. Dendrobieae and Aeridinae. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, editors. Genera Orchidacearum: Volume 6: Epidendroidea, Part 3, Oxford University Press; 2014.

19. Kocyan A, de Vogel EF, Conti E, Gravendeel B. Molecular phylogeny of Aerides (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. Mol Phylogenet Evol. 2008; 48: 422–444. https://doi.org/10.1016/j.ympev.2008.02.017 PMID: 18571438

20. Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH. Molecular phylogeny of Phalaenopsis Blume (Orchidaceae) on the basis of plastid and nuclear DNA. Plant Syst Evol. 2010; 288: 77–98.

21. Dressler RL. Phylology and classification of the orchid family. Dioscorides Press, Portland, OR. 1993.

22. Carlsward BS, Whitten WR, Williams NH, Bytebier B. Molecular phylogenetics of Vandeae (Orchidaceae) and the evolution of leaflessness. Am J Bot. 2006; 93:770–786. https://doi.org/10.3732/ajb.93.5.770 PMID: 18571438

23. Kocyan A, de Vogel EF, Conti E, Gravendeel B. Molecular phylogeny of Aerides (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. Mol Phylogenet Evol. 2008; 48: 422–444. https://doi.org/10.1016/j.ympev.2008.02.017 PMID: 18571438

24. Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH. Molecular phylogeny of Phalaenopsis Blume (Orchidaceae) on the basis of plastid and nuclear DNA. Plant Syst Evol. 2010; 288: 77–98.

25. Dressler RL. Phylology and classification of the orchid family. Dioscorides Press, Portland, OR. 1993.

26. Schuiteman A. Dendrobieae and Aeridinae. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, editors. Genera Orchidacearum: Volume 6: Epidendroidea, Part 3, Oxford University Press; 2014.

27. Kocyan A, de Vogel EF, Conti E, Gravendeel B. Molecular phylogeny of Aerides (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. Mol Phylogenet Evol. 2008; 48: 422–444. https://doi.org/10.1016/j.ympev.2008.02.017 PMID: 18571438

28. Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH. Molecular phylogeny of Phalaenopsis Blume (Orchidaceae) on the basis of plastid and nuclear DNA. Plant Syst Evol. 2010; 288: 77–98.

29. Dressler RL. Phylology and classification of the orchid family. Dioscorides Press, Portland, OR. 1993.

30. Kocyan A, de Vogel EF, Conti E, Gravendeel B. Molecular phylogeny of Aerides (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. Mol Phylogenet Evol. 2008; 48: 422–444. https://doi.org/10.1016/j.ympev.2008.02.017 PMID: 18571438
25. Kojoma M, Kurihara K, Yamada K, Sekita S, Satake M, Iida O. Genetic identification of cinnamon (Cinnamomum spp.) based on the trnL-trnF chloroplast DNA. Planta Medica. 2002; 68: 94–96. https://doi.org/10.1055/s-2002-20051 PMID: 11842343

26. Ridgway KP, Duck JM, Young JPW. Identification of roots from grass swards using PCR-RFLP and FFLP of the plastid trnL (UAA) intron. BMC Ecology. 2003; 3: 8. https://doi.org/10.1186/1472-6785-3-8 PMID: 14563214

27. Van Droogenbroeck B, Kyndt T, Romeijn-Peeters E, Scheldeman X, Romero-Motochi JP, et al. Phylogenetic analysis of the highland papayas (Vasconcellea) and allied genera (Caricaceae) using PCR-RFLP. Theor Appl Genet. 2004; 108:1473–1486. https://doi.org/10.1007/s00122-003-1575-7 PMID: 14752605

28. Peyachoknagula S, Mongkolsiriwatana C, Wannapinpong S, Huehnea PS, Srikulnatha K. Identification of native Dendrobium species in Thailand by PCR-RFLP of rDNA-ITS and chloroplast DNA. ScienceAsia. 2014; 40:113–120

29. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 1987; 4: 203–21. https://doi.org/10.1093/molbev/mal020 PMID: 3328815

30. Parani M, Rajesh K, Lakshmi M, Parducci L, Szmidt AE, Parida A. Species identification in seven small millet species using polymerase chain reaction–restriction fragment length polymorphism of trnS-psbC gene region. Genome 2001; 44: 495–499 PMID: 11444709

31. Sweet HR. Revision of the genus Phalaenopsis, Part 2. Am Orchid Soc Bull. 1968; 37: 1089–1104.

32. Fowlie JA. A note on Phalaenopsis sumatrana and Phalaenopsis cornigiana. Orchid Digest. 1969; 33: 301.

33. Christenson E A. Phalaenopsis. Timber Press, Portland Oregon, 2001; pp 330.

34. Tsai CC. Molecular Phylogeny and Biogeography of Phalaenopsis Species. In: Chen W-H, Chen HH editors. Orchid Biotechnology II. World Scientific, 2011.

35. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trnL(UAA) intron for plant DNA barcoding. Nucleic Acid Res.2007; 35: e14. https://doi.org/10.1093/nar/gkl938 PMID: 17169982