

RAPPORT D'ÉTUDE  
sur  
*CROCUS*  
(IRIDACEAE)

Critères génétiques pour  
l'identification d'une  
potentielle nouvelle  
espèce dans la section  
*versicolor*



Mai 2021

Laure Barrabé

[laure.barrabe@gmail.com](mailto:laure.barrabe@gmail.com)



## 1. Contexte

*Crocus versicolor* Ker Gawl. est une espèce connue de France et d'Italie (Figure 1), du littoral jusqu'à l'étage montagnard supérieur dans les Alpes internes et jusqu'en vallée d'Aoste, de 0 à 2300m. Des prospections récentes du Conservatoire Botanique National Alpin, notamment dans la région de Gap, ont permis d'isoler un écotype montagnard distinct des écotypes des pelouses et garrigues méditerranéennes, notamment par des critères morphologiques distincts mais peu discriminants. Ce *Crocus* de l'intérieur des terres affectionne les lisières et ourlets pré-forestiers à basse altitude et les pelouses à l'étage montagnard ou subalpin. Des observations des botanistes italiens viennent abonder l'hypothèse d'un taxon montagnard dans la région Piémont et Aoste, notamment sur substrat siliceux, différent de *C. versicolor* de climat purement méditerranéen. Il pourrait constituer une espèce nouvelle nécessitant alors d'être décrite. *Crocus crestensis* Eugène est un taxon décrit anciennement mais aujourd'hui non reconnu, inclus dans *C. versicolor* (Tison & De Foucault, 2014). Il présente des caractères morphologiques et écologiques similaires au taxon potentiellement nouveau des environs de Gap. Il est décrit des alentours de Crest (Drôme, France) mais pourrait être présent jusqu'à 1000 m d'altitude sur les moyennes montagnes de la Drôme.

Le CBNA souhaite réaliser une étude de génétique afin de déterminer à l'aide de l'outil moléculaire si ce taxon (que nous nommerons *Crocus sp. nov.*) constitue une espèce nouvelle ou correspond à des écotypes de *C. versicolor*. Nous incluons aussi des échantillons de *C. crestensis* afin de mesurer la proximité de ce dernier au nouveau taxon et à *C. versicolor*.

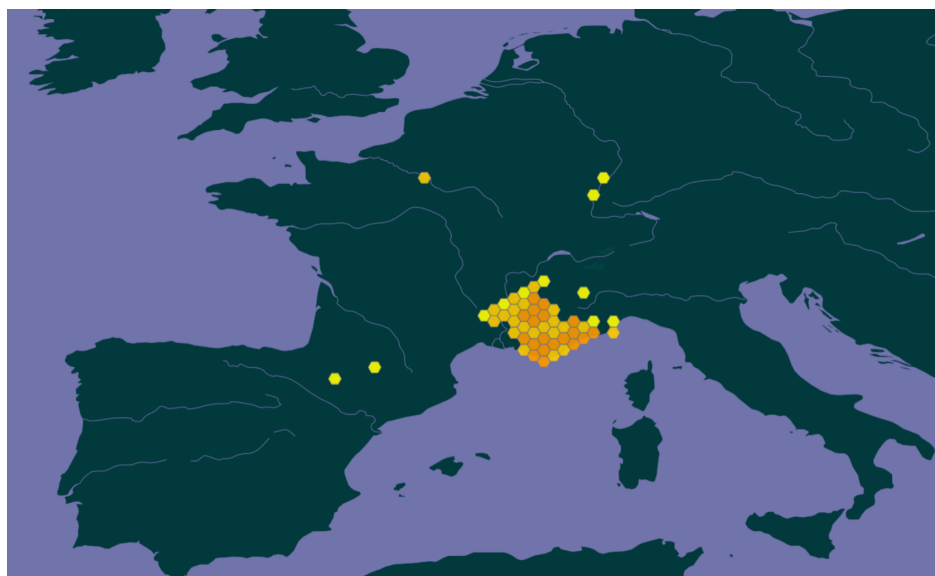


Figure 1 : Carte de distribution de *Crocus versicolor* (source GBIF : <https://www.gbif.org/species/2747502>)

Note: les points situés en Île de France, en région Sud-Ouest et en Allemagne sont des erreurs d'identification ou des plants cultivés

## 2. Méthodes

### a. Choix des marqueurs ADN

Afin de tester la monophylie<sup>1</sup> du groupe constitué par les trois taxons d'intérêt (*Crocus versicolor*, *C. crestensis* et *C. sp. nov*), nous avons sélectionné trois marqueurs ADN fréquemment utilisés dans les études phylogénétiques et taxonomiques menées sur le genre *Crocus* :

- un marqueur nucléaire ribosomique : ITS pour "Intergenic Transcribed Spacer" (Harpke et al., 2014), réputé pour son utilité dans la résolution des relations interspécifiques chez les Angiospermes. La région comprend ITS1, ITS2 et le gène rRNA 5.8S
- deux marqueurs chloroplastiques, réputés pour leur utilisation dans des études de barcoding, ainsi que pour leurs effets structurants sur les phylogénies : *ndhF* (Petersen et al., 2008), pour "chloroplast NADH dehydrogenase F", un gène codant probablement pour une protéine hydrophobe; et la région *matK / trnK* (Ford et al., 2009), respectivement pour les gènes "maturase K" et "transfer RNA".

### b. Analyses de laboratoire

Les expérimentations de laboratoire ont été confiées à la société Genoscreen (Lille, France), spécialisée dans les analyses de laboratoire génétique. Elle a donc réalisé les prestations suivantes:

- Extractions d'ADN, à partir du matériel végétal sec, fourni par le CBNA, séché en présence d'une poudre dessicante (i.e. silicagel), permettant la conservation de l'ADN.
- Synthèse des amorces PCR<sup>2</sup> nécessaires
- Amplifications d'ADN par PCR
- Séquençage des brins d'ADN

### c. Echantillonnage

Dans un premier temps, une sélection de six échantillons a été analysée. Chacun des taxons *Crocus sp. nov*, *C. crestensis* et *C. versicolor* est représenté deux fois dans cet échantillonnage: il s'agit des échantillons SA263, SA265, SA267, SA364, SA366 et SA368 (Tableau 1). Cette étape a permis de valider:

1) la qualité des extractions d'ADN, par l'évaluation à l'aide d'une électrophorèse sur gel d'agarose du niveau de dégradation de la molécule, ainsi que la quantité d'ADN obtenue, en mesurant sa concentration après extraction;

2) les amorces sélectionnées pour chacun des marqueurs, en vérifiant que celles-ci permettent d'obtenir un bon niveau d'amplification par PCR, suffisant pour procéder au séquençage; l'amplification est vérifiée et visualisée par une électrophorèse sur gel d'agarose,

3) le séquençage lui-même, par la qualité des chromatogrammes obtenus, qui permettent d'identifier avec certitude chacune des bases azotées (A, C, T, G).

Dans un second temps, une nouvelle sélection de neuf échantillons a été ajoutée pour ITS, afin d'atteindre un total de 15 échantillons pour ce marqueur. ITS étant théoriquement le marqueur le plus variable et donc le plus utile pour distinguer les taxons, il a semblé important de procéder à l'obtention de nouvelles séquences, afin de conforter les premiers résultats obtenus. Parmi ces

---

<sup>1</sup> Un groupe monophylétique contient l'espèce souche et l'ensemble de ses descendants, autrement appelé "clade"

<sup>2</sup> Une amorce est un court fragment de 15 à 25 nucléotides spécifiques de l'ADN ciblé et permettant de démarrer la polymérisation par l'enzyme Taq Polymerase utilisée dans la réaction PCR

échantillons, nous avons inclus un échantillon identifié morphologiquement comme *Crocus vernus* (L.) Hill afin d'évaluer la variabilité inter-espèces de la région ADN amplifiée avec les amorces retenues (Tableau 1).

#### d. Analyses des données de séquences

##### d.1. Nettoyage des séquences et identification taxonomique

Les séquences obtenues ont été nettoyées manuellement, assemblées et alignées grâce au logiciel Mega X (Kumar, Stecher, Li, Knyaz, and Tamura 2018). Elles ont également fait l'objet d'une identification taxonomique via l'algorithme Basic Local Alignment Search Tool, appelé communément *BLAST* (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) qui interroge la base de données internationale de recueil des séquences nucléotidiques *GenBank* (<https://www.ncbi.nlm.nih.gov/genbank>). Pour chaque séquence ADN testée, l'algorithme comparatif fournit le nom des espèces les plus proches présentes dans cette base, ainsi qu'un pourcentage d'homologie génétique avec ses dernières.

##### d.2. Variabilité des séquences

Afin d'apprécier la variabilité de chaque région d'ADN et de nos séquences, ces dernières ont été intégrées dans des alignements d'ADN incluant des espèces proches phylogénétiquement de *Crocus versicolor* (Petersen et al., 2008; Seberg & Petersen, 2009; Harpke et al., 2013), dont les séquences ont été téléchargées sous GenBank, à savoir : *C. cambessedesii* J.Gay, *C. corsicus* Vanucchi, *C. imperati* Ten., *C. ligusticus* Mariotti, *C. minimus* Redouté, *C. nudiflorus* Sm., *C. serotinus* Salisb. (que nous qualifierons de groupe interne); ainsi que des espèces un peu moins proches, à savoir : *C. banaticus* J.Gay, *C. etruscus* Parl., *C. ilvensis* Peruzzi, *C. kosaninii* Pulevic, *C. longiflorus* Raf., *C. malyi* Vis., *C. tommasinianus* Herb., *C. vernus* (L.) Herb. (que nous qualifierons de groupe externe). Les nombres de sites azotés variables (i.e. les mutations) et de sites informatifs<sup>3</sup> ont été comptabilisés pour chaque alignement. Le nombre de sites qui diffèrent entre nos séquences et celle de chaque espèce de l'alignement a également été noté.

##### d.3. Phylogénie

Un arbre phylogénétique a été reconstruit à partir des séquences ITS alignées, comprenant celles obtenues dans le cadre de notre étude ainsi que les séquences des espèces du groupe interne (incl. *Crocus versicolor*) téléchargées depuis GenBank et en utilisant *C. vernus* comme seule espèce du groupe externe. Deux méthodes de reconstruction phylogénétique ont ainsi été implémentées :

- **le maximum de vraisemblance**, à l'aide de l'algorithme phyML (Guindon & Gascuel, 2003). L'option de sélection automatique du modèle d'évolution SMS (Lefort et al., 2017) a été sélectionnée. Le modèle ainsi retenu selon le critère AIC est GTR + I + G. Les supports statistiques de branches ont été calculés selon la méthode du bootstrapping pour 1000 ré-échantillonnages (i.e. valeurs du bootstrap; BS) . Un clade est considéré comme statistiquement soutenu pour une valeur de BS supérieure ou égale à 0,9

---

<sup>3</sup> Un site azoté d'un alignement phylogénétique est considéré comme informatif lorsqu'il est partagé par au moins deux échantillons

Tableau 1. Expérimentations effectuées sur les échantillons végétaux sélectionnés de *Crocus*. Abréviations utilisées : Ech. = échantillon; Extr. ADN = Extraction d'ADN; Seq = séquençage.

			<i>ndhF</i>		ITS		<i>matK</i>	
Ech.	Taxon	Extr. ADN	PCR	Seq	PCR	Seq	PCR	Seq
SA265	<i>C. sp. nov.</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
SA267	<i>C. sp. nov.</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
SA00	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
SA367-1	<i>C. sp. nov.</i>	Non	-	-	-	-	-	-
SA369	<i>C. sp. nov.</i>	Non	-	-	-	-	-	-
SA66	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
SA49	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
SA370	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
LG45	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
SA364-1	<i>C. cretensis</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
			<i>ndhF</i>		ITS		<i>matK</i>	
Ech.	Taxon	Extr. ADN	PCR	Seq	PCR	Seq	PCR	Seq
SA365-1	<i>C. cretensis</i>	Non	-	-	-	-	-	-
SA366-1	<i>C. cretensis</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
SA263	<i>C. versicolor</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
SA264	<i>C. versicolor</i>	Oui	Non	Non	Oui	Oui	Non	Non
SA368	<i>C. versicolor</i>	Oui	Oui	Oui	Oui	Oui	Oui	Oui
AM2	<i>C. sp. nov.</i>	Oui	Non	Non	Oui	Oui	Non	Non
MB1	<i>C. sp. nov.</i>	Non	-	-	-	-	-	-
ASel1	<i>C. sp. nov.</i>	Non	-	-	-	-	-	-
ASel2	<i>C. versicolor*</i>	Non	-	-	-	-	-	-
ASel3	<i>C. versicolor*</i>	Non	-	-	-	-	-	-
ASel4	<i>C. sp. nov.</i>	Oui	Non	Non	Non	Non	Non	Non
SA372	<i>C. vernus</i>	Oui	Non	Non	Oui	Oui	Non	Non

\* l'identification taxonomique de ces échantillons n'est pas confirmée par les spécialistes botanistes

- **l'inférence bayésienne**, à l'aide du logiciel MrBayes 3.2 (Ronquist et al., 2012). Le paramétrage suivant a été utilisé : le modèle d'évolution GTR + I + G, une température de 0,2 °C, un arbre initial tiré au hasard, 1 000 000 de générations pour un échantillonnage tous les 1000 arbres. Cette analyse a été lancée trois fois indépendamment. A la fin des trois analyses, un arbre consensus de type "majority-rule" est reconstruit en utilisant tous les arbres calculés, exceptés les premiers 2500 (phase de burnin) de chaque analyse. Les supports statistiques de branches sont donnés par les probabilités a posteriori (i.e. PP), calculées durant la reconstruction de l'arbre et résumées en fin d'analyse. Un clade est considéré comme statistiquement soutenu pour une PP supérieure ou égale à 0,95.

### 3. Résultats

#### a. *ndhF*

Pour les six premiers échantillons sélectionnés, nous avons obtenu des séquences de qualité correcte sur le marqueur *ndhF* avec les amorces utilisées initialement, à savoir le couple *ndhF\_1318F* / *ndhF\_2110R*. Les séquences obtenues ont une longueur d'environ 785 paires de bases (pb). L'utilisation de BLAST permet d'identifier ces séquences comme très proches, voire identiques à des séquences attribuées au taxon *Crocus versicolor* dans la base de données GenBank, couvrant une homologie génétique (entre séquences) comprise entre 99,74 et 100% (Tableau 2). L'alignement des séquences obtenues à partir de nos échantillons permet d'identifier trois mutations non informatives.

L'alignement, incluant les espèces proches du groupe *versicolor*, dont les séquences ont été téléchargées sous Genbank, comporte 773 sites azotés, dont 56 sont variables et 22 informatifs (Tableau 3). Les groupes externe et interne comportent respectivement 22 sites variables et 11 informatifs, et 37 variables et 13 informatifs. Le groupe *versicolor* (i.e. *C. versicolor*, *C. cretensis* et *C. sp. nov.*) ne présente que trois sites variables et aucun informatif (Tableau 3). Les séquences issues des échantillons du CBNA présentent très peu de différences entre elles et avec le *C. versicolor* de Genbank (au maximum trois sites azotés variables; Tableau 4). Le groupe *versicolor* présente de nombreuses différences avec les échantillons des autres espèces du groupe interne (au minimum six sites azotés variables; Tableau 4), excepté *C. cambessedesii*.

#### b. ITS

##### b.1. Variabilité des séquences

Concernant le marqueur ITS, il a fallu procéder au test de différents couples d'amorces pour parvenir à une amplification de qualité suffisante, puis à des séquences exploitables. Le couple qui a donné les meilleurs résultats est ITS-B / ITS-D. Celui-ci permet d'amplifier la partie ITS2 de la région ciblée. Les amorces testées et utilisées sont données dans le Tableau 5. Nous avons pu ainsi obtenir des séquences d'environ 444 pb pour 14 échantillons sur 15. Les BLAST effectués sur nos séquences ITS2 permettent d'identifier ces séquences comme très proches de *Crocus versicolor* également. L'alignement des séquences obtenues à partir de nos échantillons permet d'observer une seule mutation non informative, soit une homologie génétique comprise entre 99,77 et 100% (Tableau 2).

L'alignement ITS1, incluant les espèces proches du groupe *versicolor*, comporte 274 sites azotés, dont 82 sont variables et 53 informatifs (Tableau 3). Les groupes externe et interne

Tableau 2. Résultats obtenus à partir des séquences. Abréviations utilisées : Extr. ADN = Extraction d'ADN; Seq = séquençage; ID espèce = nom de l'espèce dont la séquence est la plus similaire selon l'algorithme BLAST; % id. = pourcentage d'identité génétique (= homologie génétique); Align. = Alignement; Pos. mut. = Position des mutations

N°LABO	Taxon	Extr. ADN	ITS Nb bases = 444 - Nb mutations = 1				matK Nb bases = 933 - Nb mutations = 1				ndhF Nb bases = 785 - Nb mutations = 3			
			Blast GenBank			Align.	Blast GenBank			Align.	Blast GenBank			Align.
			ID espèce	% id.	Accession	Pos. mut.	ID espèce	% id.	Accession	Pos. mut.	ID espèce	% id.	Accession	Pos. mut.
SA265a	<i>C. sp. nov.</i>	OK	C. versicolor	99,21	HE801142.1	-	C. versicolor	100	EU496944.1	-	C. versicolor	99,74	EU110393.1	-
SA267a	<i>C. sp. nov.</i>	OK	C. versicolor	99,21	HE801142.1	-	C. versicolor	100	EU496944.1	-	C. versicolor	100	EU110393.1	-
SA00b	<i>C. sp. nov.</i>	OK	C. versicolor	98,97	HE801142.1	-	-	-	-	-	-	-	-	-
SA367-1	<i>C. sp. nov.</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
SA369	<i>C. sp. nov.</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
SA66b	<i>C. sp. nov.</i>	OK	C. versicolor	99,21	HE801142.1	-	-	-	-	-	-	-	-	-
SA49b	<i>C. sp. nov.</i>	OK	C. versicolor	99,21	HE801142.1	-	-	-	-	-	-	-	-	-
SA370b	<i>C. sp. nov.</i>	OK	C. versicolor	99,21	HE801142.1	-	-	-	-	-	-	-	-	-
LG45b	<i>C. sp. nov.</i>	OK	C. versicolor	99,23	HE801142.1	-	-	-	-	-	-	-	-	-
SA364-1a	<i>C. cretensis</i>	OK	C. versicolor	99,48	HE801142.1	-	C. versicolor	100	EU496944.1	790	C. versicolor	99,87	EU110393.1	614
SA365-1	<i>C. cretensis</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
SA366-1a	<i>C. cretensis</i>	OK	C. versicolor	99,21	HE801142.1	-	C. versicolor	100	EU496944.1	-	C. versicolor	99,74	EU110393.1	-
SA263a	<i>C. versicolor</i>	OK	C. versicolor	99,47	HE801142.1	-	C. versicolor	100	EU496944.1	-	C. versicolor	99,87	EU110393.1	421
SA264b	<i>C. versicolor</i>	OK	C. versicolor	99,23	HE801142.1	180	-	-	-	-	-	-	-	-
SA368a	<i>C. versicolor</i>	OK	C. versicolor	98,95	HE801142.1	-	C. versicolor	100	EU496944.1	-	C. versicolor	99,74	EU110393.1	431
AM2b	<i>C. sp. nov.</i>	OK	C. versicolor	99,47	HE801142.1	-	-	-	-	-	-	-	-	-
MB1	<i>C. sp. nov.</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
ASel1	<i>C. sp. nov.</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
ASel2	<i>C. versicolor*</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
ASel3	<i>C. versicolor*</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
ASel4b	<i>C. sp. nov.</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
SA372b	<i>C. vernus</i>	OK	C. vernus	94,65	AM503892.1	-	-	-	-	-	-	-	-	-

\* identification taxonomique non confirmée par les botanistes

Tableau 3. Caractéristiques des alignements incluant les séquences de *C. versicolor*, *C. sp. nov.*, *C. crestensis* (= groupe versicolor) et de leurs espèces proches. nb. seq. = nombre de séquences, nb. spp. = nombre d'espèces, nb. Sites = nombre de sites azotés, var. = nombre de sites variables, inf. = nombre de sites informatifs

	ndhF					matK					ITS1					ITS2				
	nb. seq.	nb. spp.	nb. sites	var.	inf.	nb. seq.	nb. spp.	nb. sites	var.	inf.	nb. seq.	nb. spp.	nb. sites	var.	inf.	nb. seq.	nb. spp.	nb. sites	var.	inf.
alignement complet	46	16-19	773	56	22	29	15-18	925	57	13	52	16	274	82	53	66	16-19	405	69	51
groupe externe	28	8		22	11	11	7		36	8	28	8		47	23	29	8		23	11
groupe interne	18	8-11		37	13	18	8-11		17	4	24	8		61	28	37	8-11		60	37
groupe interne excepté groupe <i>versicolor</i>	11	7		26	9	11	7		15	3	23	7		56	26	23	7		54	31
groupe <i>versicolor</i>	7	1-3		3	0	7	1-3		2	0	1	1		NA	NA	14	1-3		3	0



Tableau 4. Nombre de sites variables entre les différents taxons de la section versicolor, pour chacune des régions ADN étudiées.

ITS1							
	cambessedesii	corsicus	imperati	ligusticus	minimus	nudiflorus	serotinus
versicolor GenBank	13	27	20	30	26	22	22

ITS2										
	cambessedesii	corsicus	imperati	ligusticus	minimus	nudiflorus	serotinus	versicolor CBNA	crestensis CBNA	sp nov CBNA
versicolor GenBank	13	16	14	22	16	22	20	3	2	2
versicolor CBNA	16	19	17	25	19	25	23	-	1	1
crestensis CBNA	15	18	16	24	18	24	22	1	-	0
sp nov CBNA	15	18	16	24	18	24	22	1	0	-

<i>matK</i>										
	cambessedesii	corsicus	imperati	ligusticus	minimus	nudiflorus	serotinus	versicolor CBNA	crestensis CBNA	sp nov CBNA
versicolor GenBank	1	1	7	3	3	1	3	1	2	1
versicolor CBNA	2	2	8	4	4	2	4	-	1	0
crestensis CBNA	3	3	9	5	5	3	5	1	-	1
sp nov CBNA	2	2	8	4	4	2	4	0	1	-

<i>ndhF</i>										
	cambessedesii	corsicus	imperati	ligusticus	minimus	nudiflorus	serotinus	versicolor CBNA	crestensis CBNA	sp nov CBNA
versicolor GenBank	1	8	19	9	7	6	9	2	1	0
versicolor CBNA	3	10	28	11	9	8	11	-	3	2
crestensis CBNA	2	9	28	10	8	7	10	3	-	1
sp nov CBNA	1	8	27	9	7	6	9	2	1	-

comportent respectivement 47 sites variables et 23 informatifs, et 61 variables et 28 informatifs. Le *Crocus versicolor* de Genbank présente de nombreuses différences avec les échantillons des autres espèces du groupe interne (au minimum 13 sites azotés variables; Tableau 4).

L'alignement ITS2, incluant les espèces proches du groupe versicolor, comporte 405 sites azotés, dont 69 sont variables et 51 informatifs (Tableau 3). Les groupes externe et interne comportent respectivement 23 sites variables et 11 informatifs, et 60 variables et 37 informatifs. Le groupe *versicolor* ne présente que trois sites variables et aucun informatif. Les séquences issues des échantillons du CBNA présentent très peu de différences entre elles et avec le *C. versicolor* de Genbank (au maximum trois sites azotés variables; Tableau 4). Le groupe *versicolor* présente beaucoup de différences avec les échantillons des autres espèces du groupe interne (au minimum 13 sites azotés variables; Tableau 4).

## b.2. Phylogénie

Les reconstructions phylogénétiques obtenues à partir des séquences ITS selon le maximum de vraisemblance et l'inférence bayésienne donnent des résultats similaires. Nous ne présenterons ainsi que l'arbre consensus bayésien (Figure 1). Ce dernier montre que les échantillons de notre étude forment un clade<sup>1</sup> très fortement soutenu avec *Crocus versicolor* (PP de 1). Nos échantillons forment eux-mêmes un clade unique, également fortement soutenu (PP de 1). Le clade soutenu (PP = 1) formé par les échantillons de *C. cambessedesii* est placé comme groupe-frère<sup>4</sup> de nos échantillons et de *C. versicolor* (PP = 1; Figure 1).

Tableau 5. Couples d'amorces testées et travaux effectués pour les marqueurs ITS et *matK*. Soulignés et en gras: les couples retenus pour cette étude

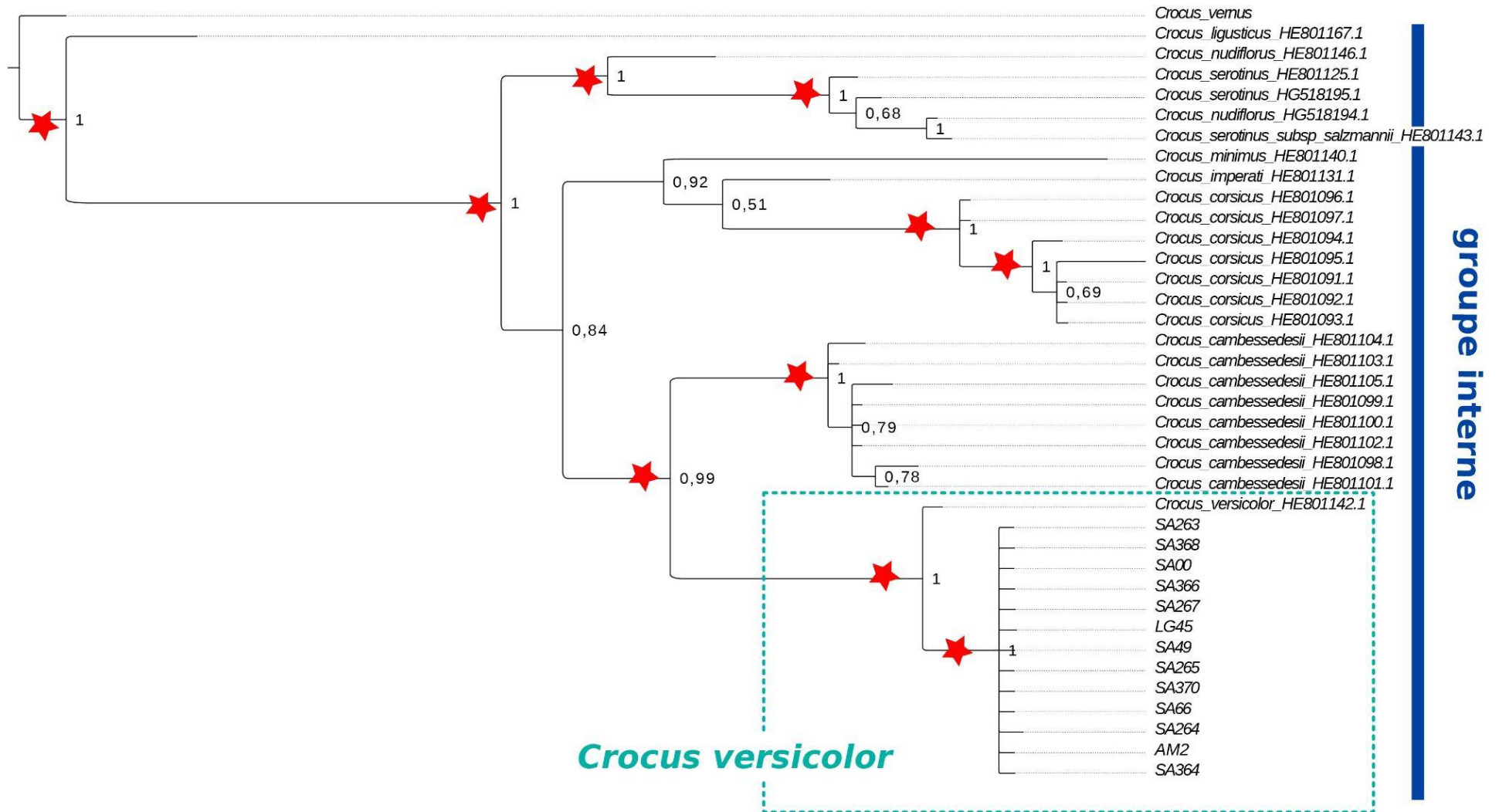
Marqueur	Couple	PCR	Séquençage
ITS	ITS-A / ITS-B	Oui	Non
	ITS-C / ITS-D	Oui	Non
	<u>ITS-B / ITS-D</u>	<u>Oui</u>	<u>Oui</u>
<i>matK</i>	matK2.1a / matK-5R	Oui	Non
	matK-X / matK-3.2	Oui	Non
	<u>matK-X / matK-5R</u>	<u>Oui</u>	<u>Oui</u>

## c. *matK*

Concernant le marqueur *matK*, il a fallu procéder à différents tests d'amorces en PCR puis en séquençage. Les amorces testées et utilisées sont données dans le Tableau 5. Le couple d'amorces qui a permis d'obtenir des séquences exploitables est matK-X / matK-5R. Nous avons ainsi pu obtenir des séquences d'environ 933 pb pour les six premiers échantillons sélectionnés. L'alignement des séquences permet d'observer une seule mutation non informative. Les BLAST effectués sur nos séquences *matK* rejoignent les résultats obtenus sur les deux précédents marqueurs, montrant une

<sup>4</sup> des clades formant ensemble un clade entier sont dits **groupes frères** les uns des autres

Figure 1. Arbre phylogénétique bayésien obtenu à partir des séquences ITS. Le code inscrit en fin de nom d'une séquence correspond au numéro d'accension GenBank. Seuls les supports de nœuds dont la probabilité a posteriori est supérieure à 0,5 sont indiqués. Les étoiles rouges désignent les nœuds dont la PP est supérieure à 0,95.



homologie génétique de 100% avec *Crocus versicolor* (Tableau 2). L'alignement, incluant les espèces proches, comporte 925 sites azotés, dont 57 sont variables et 13 informatifs (Tableau 3). Les groupes externe et interne comportent respectivement 36 sites variables et huit informatifs, et 17 variables et quatre informatifs. Le groupe *versicolor* ne présente que deux sites variables et aucun informatif (Tableau 3). Les séquences issues des échantillons CBNA présentent très peu de différences entre elles et avec le *C. versicolor* de Genbank (au maximum deux sites azotés variables; Tableau 4). Le groupe *versicolor* présente peu de différences avec les échantillons des autres espèces du groupe interne (au maximum sept sites azotés variables; Tableau 4).

Au vu de l'ensemble des résultats portant sur les trois régions ADN, il a été décidé de ne pas poursuivre le séquençage sur les échantillons restants.

#### **4. Discussion et conclusions**

##### **a. Identité génétique**

Les marqueurs génétiques employés dans notre étude sont parmi les plus discriminants pour *Crocus*, et utilisés dans des études récentes portant sur le genre afin d'identifier des espèces nouvelles (Erol et al., 2015 ; Miljković et al., 2016) et de reconstruire des phylogénies (Petersen et al., 2008; Seberg & Petersen, 2009; Harpke et al., 2013, Nemati et al., 2018).

Les résultats obtenus sur ces trois régions d'ADN montrent une très faible variabilité génétique entre les trois taxons d'intérêt, définis initialement sur une base morphologique : pour chacune des régions étudiées, l'homologie génétique entre les séquences est supérieure ou égale à 99,7% (Tableau 2), bien trop élevée pour considérer les taxons étudiés comme différents. En outre les séquences de ces trois taxons présentent entre elles très peu de bases azotées différentes (Tableaux 3 et 4), alors que comparativement avec les autres espèces proches (i.e. du groupe interne), ces différences sont beaucoup plus élevées, particulièrement pour les régions ITS2 et *ndhF* (Tableaux 3 et 4). Ce deuxième résultat conforte l'hypothèse que les taxons *Crocus crestensis* et *C. sp. nov.* appartiennent à la même entité taxonomique que *C. versicolor*.

L'arbre phylogénétique reconstruit confirme l'appartenance de nos échantillons à un clade incluant *Crocus versicolor*. Le regroupement (fortement soutenu; PP = 1) de nos échantillons placés comme groupe-frère de cette espèce est probablement la conséquence du séquençage partiel de nos séquences (ca. 440 pb vs. 630 pb pour les séquences téléchargées sous genbank).

##### **b. Biais de l'étude**

###### **b.1. Variabilité des marqueurs chloroplastiques**

Utilisées seules, les régions chloroplastiques *matK* et *ndhF* ne sont pas suffisamment variables pour distinguer les espèces chez le genre *Crocus* de manière certaine (Petersen et al., 2008, Seberg & Petersen, 2009). Nous avons décidé de les intégrer à notre étude pour leurs effets phylogénétiques structurants et pour cumuler leurs variabilités génétiques. Le faible niveau de divergence génétique sur ces marqueurs vient conforter les résultats obtenus sur ITS.

###### **b.2. Caractéristiques génétiques d'ITS**

Le séquençage du marqueur ITS n'a permis d'obtenir que la partie ITS2 de la région visée, soit 444 paires de bases environ. Les séquences de la région ITS1 auraient permis de conforter les résultats obtenus, mais ni le couple d'amorces ITS-A/ITS-B ni le couple ITS-C/ITS-D n'ont pu permettre d'obtenir des séquences de bonne qualité. Toutefois, l'alignement de séquences ITS de *Crocus* (séquences téléchargées depuis la base GenBank) ne semble pas montrer de différence de variabilité entre les parties ITS1 et ITS2. Pour les espèces du groupe interne (i.e. espèces proches de *C. versicolor*), on comptabilise environ le même nombre de mutations entre ITS1 et ITS2 (ca. 60; Tableau 3), cette dernière possédant de surcroît plus de sites informatifs (37 vs. 28; Tableau 3). L'utilisation de la région ITS2 seule peut donc être amplement suffisante pour discriminer des espèces proches.

ITS est une région ADN connue pour parfois subir le phénomène de paralogie. Chez certaines espèces, elle peut être présente en plusieurs copies (i.e. les pseudogènes) pouvant évoluer et muter indépendamment les unes des autres, et pouvant perturber les reconstructions phylogénétiques. Ce phénomène peut ainsi entraîner une mauvaise interprétation des résultats. Toutefois ce phénomène est contrebalancé par un mécanisme de correction naturelle des copies d'ADN (i.e. l'évolution concertée). ITS est un marqueur très utilisé chez le genre *Crocus* et suffisamment variable pour permettre de séparer les espèces dans ce genre (Erol et al., 2017 ; Miljković et al., 2016 ; Erol et al., 2015 ; Kerndorff et al., 2013). En outre, des études récentes ont montré que l'existence de pseudogènes chez certaines espèces de *Crocus*, ne faussait pas les résultats taxonomiques et phylogénétiques (Harpke et al., 2008a et 2008b ; Erol et al., 2015 ; Miljković et al., 2016).

### **b.3. Les données Genbank**

Genbank est peu fourni en échantillons de *Crocus versicolor*, limitant les comparaisons. Seuls deux échantillons ont leurs séquences ITS, *ndhF* ou *matK* disponibles dans la base de données. Ces deux échantillons sont identifiés sous les identifiants "C1558" et "GAT:7270". Les séquences correspondantes ont été utilisées dans les publications suivantes:


- séquence ITS - Harpke et al. 2013
- séquence *ndhF* - Petersen et al. 2008
- séquence *matK* - Seberg & Petersen 2009

Ces deux échantillons n'ont pas été collectés dans le même pays (respectivement France et en Italie), ce qui peut être critiquable en termes de fiabilité taxonomique, si des incertitudes pèsent sur la distribution géographique de l'espèce.

Dans tous les cas, les nouvelles séquences, obtenues dans le cadre de la présente étude constituent une contribution importante pour la connaissance du genre *Crocus* car elles devraient enrichir considérablement la base de données Genbank existante.

### **c. Perspectives**

Au vu des résultats, il nous apparaît peu probable que la forte homologie de séquences obtenue sur les trois régions ciblées ne soit pas représentative d'une réalité taxonomique, à savoir que les taxons *Crocus crestensis* et *C. sp. nov.* appartiennent à la même entité spécifique que *C. versicolor*.




Il serait plus vraisemblable que les trois entités taxonomiques, définies sur une base morphologique, soient des écotypes inféodés à certains milieux ou certaines zones géographiques. Pour explorer cette possibilité, plusieurs méthodes peuvent être mises en œuvre. Il peut être envisagé d'utiliser des marqueurs génétiques capables de décrire la diversité au niveau intra-spécifique (populationnel), au sein de l'espèce *Crocus versicolor*, en incluant les populations de *C. cretensis* et *C. sp. nov.* Pour cela, des marqueurs tels que les microsatellites ou les RADseq permettraient d'atteindre un niveau de résolution suffisant.

Il peut être aussi envisagé de réaliser une étude génomique, de type séquençage complet, soit profond ("whole-genome sequencing"), soit de surface ("genome skimming"). Néanmoins des méthodes d'enrichissement (séquençage partiel) peuvent s'avérer suffisantes, car séquençant plusieurs centaines de loci en ciblant des régions conservées du génome, telles que "Anchored Hybrid Enrichment" (AHE) ou "Ultraconserved Elements" (UCE) (Young & Gillung, 2020).

## Bibliographie

- Erol, O., Harpke, D. and Yildirim, H. (2015). A new *Crocus* l. (Iridaceae) species from SE Turkey, based on morphological and molecular data. *Phytotaxa*, 239(3): 223–232. <https://doi.org/10.11646/phytotaxa.239.3.3>
- Erol, O., Harpke, D. and Ciftci, A. (2017). *Crocus heilbronniorum*, a new Turkish species of Series *Lycioteuri* (Iridaceae). *Phytotaxa*. <https://doi.org/10.11646/phytotaxa.298.2.7>
- Eugène, R.P. (1868). Note sur une nouvelle espèce de crocus. *Bulletin de la Société Botanique de France*, 15(9): 190–191.
- Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S. B., Cowan, R. S., Chase, M.W. and Wilkinson, M.J. (2009). Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society*, 159(1): 1–11. <https://doi.org/10.1111/j.1095-8339.2008.00938.x>
- Guindon, S. & Gascuel, O. (2003) PhyML : A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 52(5):696-704. <https://doi.org/10.1080/10635150390235520>
- Harpke, D. and Peterson, A. (2008) 5.8S motifs for the identification of pseudogenic ITS regions. *Botany*. 86(3): 300-305. <https://doi.org/10.1139/B07-134>
- Harpke, D. and Peterson, A. (2008) Extensive 5.8S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer regions. *J Plant Res* 121, 261–270. <https://doi.org/10.1007/s10265-008-0156-x>
- Harpke, D., Meng, S., Rutten, T., Kerndorff, H., and Blattner, F.R. (2013). Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: Ancient hybridization and chromosome number evolution. *Molecular Phylogenetics and Evolution*, 66(3): 617–627. <https://doi.org/10.1016/j.ympev.2012.10.007>
- Lefort, V., Longueville, J-E., and Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, Volume 34, Issue 9, September 2017, Pages 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Kerndorff, H., Pasche, E., Blattner, F.R. and Harpke D. (2013). A New Species of *Crocus* (Liliiflorae, Iridaceae) from Turkey. *STAFPIA* 99: 141-144.
- Miljković, M., Ranđelović, V. and Harpke, D. (2016). A new species of *Crocus* (Iridaceae) from southern Albania (SW Balkan Peninsula). *Phytotaxa*, 265(1): 39–49. <https://doi.org/10.11646/phytotaxa.265.1.3>



Nemati, Z., Blattner, F.R., Kerndorff, H., Erol, O., & Harpke, D. (2018). Phylogeny of the saffron-crocus species group, *Crocus* series *Crocus* (Iridaceae). *Molecular Phylogenetics and Evolution*, 127: 891–897. <https://doi.org/10.1016/j.ympev.2018.06.036>

Petersen, G., Seberg, O., Thorsøe, S., Jørgensen, T. and Mathew, B. (2008). A phylogeny of the genus *Crocus* (Iridaceae) based on sequence data from five plastid regions. *Taxon*, 57(2): 487–499

Ronquist, F., M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, and J.P. Huelsenbeck. 2012. MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Systematic Biology*. 61:539-54. <https://doi.org/10.1093/sysbio/sys029>

Seberg, O., & Petersen, G. (2009). How many loci does it take to DNA barcode a crocus? *PLoS ONE*, 4(2): 2–7. <https://doi.org/10.1371/journal.pone.0004598>

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547-1549. <https://doi.org/10.1093/molbev/msy096>

Tison, J.-M. and De Foucault, B. (coords) (2014). *Flora Gallica - Flore de France*, ed. Biotope (Mèze), 1196 p.

Young, A.D. & Gillung, J.P. (2020). Phylogenomics - principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology* 45: 225-247. <https://doi.org/10.1111/syen.12406>