

Identification of Domesticated Silkworm Varieties Using a Whole Genome Single Nucleotide Polymorphisms-based Decision Tree

Jong Woo Park^{1†}, Jeong Sun Park^{2†}, Chan Young Jeong¹, Hyeok Gyu Kwon¹, Sang Kuk Kang¹, Seong-Wan Kim¹, Nam-Suk Kim¹, Kee Young Kim¹ and Iksoo Kim^{2*}

¹Department of Agricultural Biology, National Academy of Agricultural Science, Rural Development Administration, Wanju Gun, Jeollabuk-do 55365, Korea

²Department of Applied Biology, College of Agriculture & Life Sciences, Chonnam National University, Gwangju 61186, Korea

Received October 5, 2022 /Revised November 21, 2022 /Accepted November 21, 2022

Silkworms, which have recently shown promise as functional health foods, show functional differences between varieties; therefore, the need for variety identification is emerging. In this study, we analyzed the whole silkworm genome to identify 10 unique silkworm varieties (Baekhwang, Baekok, Daebaek, Daebak, Daehwang, Goldensilk, Hansaeng, Joohwang, Kumkang, and Kumok) using single nucleotide polymorphisms (SNP) present in the genome as biomarkers. In addition, nine SNPs were selected to discriminate between varieties by selecting SNPs specific to each variety. We subsequently created a decision tree capable of cross-verifying each variety and classifying the varieties through sequential analysis. Restriction fragment length polymorphism (RFLP) was used for SNP867 and SNP9183 to differentiate between the varieties of Daehwang and Goldensilk and between Kumkang and Daebak, respectively. A tetra-primer amplification refractory (T-ARMS) mutation was used to analyze the remaining SNPs. As a result, we could isolate the same group or select an individual variety using the nine unique SNPs from SNP780 to SNP9183. Furthermore, nucleotide sequence analysis for the region confirmed that the alleles were identical. In conclusion, our results show that combining SNP analysis of the whole silkworm genome with the decision tree is of high value as a discriminative marker for classifying silkworm varieties.

Key words : Markers, silkworm varieties, SNP, T-ARMS

서 론

누에(*Bombyx mori*)는 곤충강 나비목 누에나방과에 속하는 곤충으로서, 천잠(*Antheraea yamamai*), 작잠(*Antheraea pernyi*) 등과 같은 야생 누에와 비교하여 집누에라고도 한다[6]. 누에는 뽕밭에 야생하는 멧누에(*Bombyx mandarina*)와 교배하여 생식력이 있는 자손을 만들 수 있는 근연종이며, 누에 품종은 그 원산지에 따라 일본종, 중국종, 유럽종, 재래종으로 나눌 수 있고, 휴면의 형태에 따라 여러 세대 경과 후 휴면에 들어가는 다화성(multivoltine) 또는 1년에 1세대 만 발생하는 일화성(univoltine)으로 분류하기도 한다[6]. 한국 농가에서 사육이 가능한 누에 보

급품종은 일본종계와 중국종계의 육성 원종을 교배하여 만들어진 1대 교잡종으로[10], 총 23종이 장려품종으로 지정되어 보급되고 있다. 이와 같은 장려품종은 무늬의 유무에 따라 누에의 암수를 감별할 수 있는 한성품종(sex-limited race)인 한생잠 및 도담누에 등과 인공사료에 적합한 품종으로 백옥잠 및 금옥잠 등 용도에 따라 육성된 다양한 품종이 존재한다[8]. 또한 고치색에 따른 분류가 가능하며, 흰색 고치를 짓는 품종으로 금옥잠, 대박잠, 백옥잠 금강잠이 있으며, 황색고치를 짓는 품종으로는 대황잠, 백황잠 및 골든실크가 있다. 그 밖에 녹색 고치를 짓는 품종으로 연녹잠, 그리고 주황색 고치를 짓는 품종으로 주황잠 등이 보급되어 사육되고있으나, 그중 백옥잠과 골든실크가 전체 누에 생산량의 98% 이상을 차지할 정도로 높은 비중을 차지하고있다[10-13].

과거에는 섬유산업에 이용되어 오던 누에가 최근 알츠하이머 예방, 혈당 강하 및 콜레스테롤 저하 등의 다양한 기능성을 나타낸다고 알려지며 건강 기능성식품의 원료로서 주목받게 되면서[1, 9, 24], 고부가가치 상품 개발 및 최근 소비트렌드에 맞는 상품의 개발이 이루어지고 있다 [9]. 또한 품종 별 기능성의 차이에 대한 심도 있는 연구가

[†]Authors contributed equally.

*Corresponding author

Tel : +82-62-238-2913, Fax : +82-63-238-3832

E-mail : ikkim81@chonnam.ac.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

진행되며 혈당강하 및 간 기능 개선에 특화된 품종으로 백옥잠 및 금옥잠[23], 인지능력 개선과 관련된 기능성 우수품종으로 골든실크[9]가, 탈모예방에 관한 기능성으로 주황잠, 피부노화 및 항산화와 관련된 품종으로 연녹잠 [20] 등이 주목받게 됨에 따라 누에를 용도별 건강 기능성 식품으로 발전시키기 위해서는 품종 판별을 포함하는 품질관리 기술이 필요한 실정이다.

최근 생물체의 게놈 분석 기술이 비약적으로 발전함에 따라 다양한 생물에서 DNA 다형성에 기반한 유전적 마커를 이용한 종의 기원, 품종 형성, 유전적 다양성 및 다른 품종과의 유연관계를 분석하고자 하는 연구들이 활발히 진행되고 있다[5]. 이와 같이 식물, 동물 등의 연구 및 육종 프로젝트에서 광범위하게 이용되고 있는 분자마커로는 제한효소 부위 변이에 따른 길이 차이를 분석하는 Restriction Fragment Length Polymorphism (RFLP) 및 단일 염기에서 발생된 변이인 Single Nucleotide Polymorphism (SNP) 등이 다양하게 이용되고 있으며, 그중 SNPs를 이용한 분자마커는 대립형질의 변이와 초고처리량 및 자동화 기술에 적용될 수 있는 장점이 있어, 다른 분자마커들에 비해 가장 널리 이용되고 있다[25].

따라서, 본 연구에서는 장려품종 누에 10종의 전장유전체 서열을 분석하고, 각 품종에 특이적인 2종의 SNP를 선별하여 이를 이용한 decision tree 작성 및 순차 진단을 통한 주요 누에 장려 품종에 대한 판별을 실시하고 마커로서 사용하기 위한 방법을 탐색하고자 하였다.

재료 및 방법

실험재료

본 실험에 사용된 장려품종 누에는 품종 육성 용 원종 교배종인 Daehwang (Jam 323×324), Joohwang (Jam 325×326), Goldensilk (Jam 311×312), Daebaek (Jam 157×162), Hansaeng (Jam 153×154), Daebak (Jam 155×156), Kumkang (Jam 159×160), Kumok (Jam 125×140), Baekok (Jam 123×124), Baekhwang (Jam 327×328)을 이용하였으며, 누에 사육은 농촌진흥청 농업생물부의 표준 사육기준(온도, 24~27°C, 습도, 70~90%)에 준하여 신선한 뽕잎을 급여하고 25°C, 16L8D의 조건에서 사육하였다.

전장 유전체 서열 분석 및 SNP 선별

사육된 누에로부터 genomic DNA를 추출하기 위하여 10개 품종에 대하여 3령기 유충을 각각 10마리씩 준비 후, 유충의 가슴 조직으로부터 Wizard™ Genomic DNA Purification Kit (Promega, USA)를 사용하여 genomic DNA를 추출하였다. 추출된 핵산은 전기영동 및 Dropsense96 (Trinean, Gentbrugge, Belgium)와 Picogreen (Invitrogen, Oregon, USA)을 이용한 흡광도 분석을 통하여 서열 분석

적합도를 분석하고, 각 누에 품종에 대한 전장유전체 염기서열을 확보하기 위하여 QC가 완료된 DNA와 MGIEasy DNA library prep Kit (BGI, Shenzhen, China)를 이용하여 sequencing library를 구축하였다. 구축된 library는 MGISEQ-2000 (BGI, Shenzhen, China)을 이용하여 read length 150 bp paired-end 방식으로 데이터 생산을 진행하고, 생산된 데이터에 대해 index sequence를 이용한 de-multiplexing을 수행하여 누에 10개 품종 100개체의 개별 시료에 대한 서열 데이터를 확보하였다. 전장 유전체 기반의 품종 별 SNP 선별을 위하여 기 분석된 고품질의 집누에 유전체를 참조유전체로 이용하여[16], re-sequencing을 통해 개별 시료의 변이정보를 확보하였다. 먼저 생산된 데이터는 데이터 생산과정에서 발생할 수 있는 저품질 서열과 sequencing artifact의 제거를 위해 Trimmomatic [2]을 사용하여 read 말단의 quality score 20 미만의 염기 및 MGISEQ의 adapter 서열을 제거하였다. QC가 완료된 paired-end read는 BWA-MEM2 [26]를 이용하여 Indexing된 집누에 참조 유전체 서열에 mapping되었으며, mapping 결과는 Samtools [18]의 rmdup을 이용하여 데이터 생산과정에서 생성된 PCR duplicated read를 제거하였다. Variant calling은 Samtools를 이용하여 진행되었으며 최초 확보된 총 변이는 VCFtools [4]를 이용하여 후속 분석을 위한 필터링을 진행하였다. 필터링은 먼저 InDel을 제거하여 SNP를 추출하고 variant quality score가 30 미만이고 전체 시료에서 genotyping 정보가 확보되지 않아 missing genotype을 가지는 변이를 필터링하였다. 또한, 하나의 locus에 여러 개의 Alternative allele이 발생하는 Multi-allelic region의 경우 Sequencing Error와 및 후속 분석에서의 복잡성을 감안하여 분석에서 제외하였으며, decision tree를 이용할 후속분석에서의 분류 모델의 정확도 확보를 위하여 해당 locus에 존재하는 genotype이 hetero type이 없이 homo type만이 존재하는 variant loci를 추출하여 후속 분석에 사용하였다. Decision tree를 이용한 분류 모델 구축을 위해 필터링을 통해 확보된 변이 정보는 Weka [6]의 decision tree classifier인 J48 [24]모듈을 이용하여 10-fold cross-validation 기준으로 100% prediction accuracy를 보이는 마커들을 선별하였다.

판별마커 영역에 대한 프라이머 제작

SNP 검증을 위하여 Medrano and Oliveira [19]의 방법에 따른 Tetra-primer Amplification Refractory Mutation System (T-ARMS PCR)과 Parties *et al.* [22]에 따른 RFLP 방법을 이용하였다. Tetra-ARMS PCR은 4개 primer를 이용하여 한번의 PCR 수행 후 전기영동 결과를 바탕으로 SNP를 판별하는 방식으로, 대립유전자 영역을 공통적으로 증폭하기 위한 외부 primer 쌍(Outer primer; OF and IR)과 SNP 유전자형 판별하기 위해 각각의 대립유전자에 특이적으로 반응하여 증폭하는 내부 primer (Inner primer; IF and

IR) 쌍으로 제작하고 각 SNP에 대한 primer는 Table 1에 나타내었다. Tetra-ARMS PCR 방법으로 확인이 어려운 SNP에 대해서는 RFLP 방법을 이용하기 위하여, SNP를 중심으로 증폭된 유전자에 제한효소 처리에 따른 크기 차이가 표현 될 수 있도록 primer를 제작하고 Table 2에

나타내었다.

PCR을 통한 SNP 영역 증폭 및 염기서열 검증
 준비된 T-ARMS PCR 및 RFLP 검증을 위한 Table 1의 프라이머와 서열분석을 위하여 추출한 genomic DNA를

Table 1. Sequence of T-ARMS-PCR primers used for the identification of silkworm varieties

SNP	SNP site	Primer sequence (5'→3')	Size (bp)	Varieties
SNP780	(A/T)	SNP780-OF: AATRCGCCTAAATAAGTTTCAC	265	
		SNP780-OR2: AAGCCTTTCGGTTACGTCATG		
		SNP780-IF3: CGTCTGAATTAATATTACAACA <u>A</u>	135	Etc
		SNP780-IR2: GAACACGAAACACGCATTAG <u>CCA</u>	175	DH & GS
SNP867	(A/G)	SNP867-OF: TGAGATCTTCTTCAGTTTGCAG	501	
		SNP867-OR: CTYCACCTAATCGTTGGTATC		
		SNP867-IF: CGACAAAGTTAATAAGAA <u>TGA</u>	376	DH
		SNP867-IR: ATGAACTAAGATTAATGAA <u>AGC</u>	167	GS
SNP1271	(T/C)	SNP1271-OF: TACGAGTATTGCCTTGAACG	414	
		SNP1271-OR: GTAATTGCTAAGTATTTTATGC		
		SNP1271-IF: CGAAAGGTTAGCTTTTTCG <u>IGT</u>	278	Etc
		SNP1271-IR: ATAACTATTCTACGTAATGCT <u>TAG</u>	180	KO & BH
SNP427	(T/C)	SNP427-OF2: ATACCAAGAGCTGTTTCATACG	319	
		SNP427-OR: ACCGAAAATAATGTTAATGACG		
		SNP427-IF: CGAGTATATGTAGTAGGTAT <u>TAT</u>	146	KO
		SNP427-IR: ATTTTAAAATATTTTACAAAAT <u>GAG</u>	220	BH
SNP4165	(T/C)	SNP4165-OF: GGTGAAGTTGCAAAAATGTTGC	487	
		SNP4165-OR: ATTTACTAAGAGTTTCAGTGGAC		
		SNP4165-IF3: CAGATTTAAAATGTGAAAT <u>CAT</u>	138	JH
		SNP4165-IR2: ATTAACAGTTAATAGACCAT <u>TCG</u>	392	Etc
SNP4381	(G/A)	SNP4381-OF: TCGTAAAGCGCATCACATTTCG	391	
		SNP4381-OR: AYGICTACRTCGCGTCATC		
		SNP4381-IF: GTCAAAAGTGGATCGGACTAT <u>TG</u>	164	Etc
		SNP4381-IR: CGTTTACAACATAATAGAG <u>CGT</u>	271	HS
SNP4545	(G/T)	SNP4545-OF2: GCACGTATGTAATTTTGATTG	262	
		SNP4545-OR: AATAAATTACAGTTTCMGAAGC		
		SNP4545-IF3: CCAACCGGCAGATTCC <u>TAG</u>	106	BO
		SNP4545-IR2: CAGGATGGTAATATTATATCC <u>AA</u>	197	KK & DBK

*Bpold letters: SNP site for the identification of varieties.

**Underlined bold letters: additional substitution to increase amplification efficiency.

DB, DaeBak; GS, GoldenSilk; KK, KumKang; DH, DaeHwang; BH, BaekHwang; BO, BaekOk; DBK, DaeBaek; HS, HanSaeng; JH, JooHwang; KO, KumOk.

Table 2. Sequence of T-ARMS-PCR primers used for the identification of silkworm varieties

SNP	SNP site	Primer sequence (5'→3')	Restriction enzyme	Varieties/ Amplicon (bp)
SNP 1169	(C/T)	SNP1169-F: TTTTAAGCCTTATTTAATTTAGTC	<i>Ahd</i> I	DB/108+634+238
		SNP1169-R: TGACGAGTGTGGACAACCTCAG	GACNNN/NGTC	Etc/108+872
SNP 9183	(T/A)	SNP9183-F: TTTATTARATGGTTAAACTAGC	<i>Ssp</i> I	KK/109+434
		SNP9183-R: TACTATGGTAACATACCGTC	AAT/ATT	DBK/109+268+167

*Bold letters: SNP site for the identification of varieties.

DB, DaeBak; GS, GoldenSilk; KK, KumKang; DH, DaeHwang; BH, BaekHwang; BO, BaekOk; DBK, DaeBaek; HS, HanSaeng; JH, JooHwang; KO, KumOk

이용하여 PCR을 수행하였다. 각 영역의 증폭은 Taq DNA polymerase (Takara, Japan)와 Thermal cycler (Takara, Japan)를 이용하여 94°C에서 2분간 초기 변성 후, 94°C에서 1분, 50°C에서 1분 및 72°C에서 1분 주기를 30회 반복하고, 최종적으로 72°C에서 4분간 각 SNP 영역을 증폭하였다. Tetra-ARMS PCR primer를 이용한 증폭 산물의 경우 바로 2% agarose gels를 이용하여 전기영동하고, RFLP 분석을 위한 증폭 산물의 중 SNP867에 대해서는 *Ahd I* 효소와 SNP9183에 대해서는 *Ssp I* 효소를 한시간 동안 처리 후 2% agarose gels에 전기영동하여 밴드의 크기를 비교하였다. 각 증폭산물에 대한 대립유전자 서열 검증을 위하여 외부 primer 쌍에 의해 증폭된 밴드를 pGem-T easy vector system (Promega, USA)에 클로닝 후 universal primer를 이용하여 서열 분석을 실시하였다.

결과 및 고찰

전장 유전체 서열 분석

누에 품종 구분 마커 개발을 위하여, 백옥잠, 대황잠, 주황잠, 골든실크, 대백잠, 한생잠, 대박누에, 금강잠, 금옥잠, 백황잠 10개 품종을 대상으로 각 품종별 10개체씩 DNA를 추출하고 농도 및 순도를 분석한 결과 NGS 서열 분석에 충분한 DNA가 확보되었다. 염기서열 분석을 위하여 데이터를 생산한 결과, 총 587,123,548,200 bp의 염기서열이 분석되었고, 품질점수는 86.85%가 Q30 이상으로 99.9%의 정확성을 나타냈다. 생산된 데이터에 대해 Index sequence를 이용한 De-multiplexing을 수행하여 누에 10개 품종 100개체의 개별 시료에 대한 데이터를 확보한 결과, 100개체의 시료 중 최소 데이터 생산량은 약 4.3 Gb 였으며, 이러한 결과는 보편적인 곤충의 게놈의 크기를 1 Gb 미만으로 가정했을 때, 각 시료에서 일반적인 곤충의 약 4배 이상의 서열데이터가 확보된 것 판단된다.

SNP 선발 및 decision tree 작성

품종 특이적인 SNP 선발을 위하여 염기서열분석 결과와 기 분석된 집누에의 참조유전체 서열정보를 바탕으로

Samtools를 이용한 변이추출을 진행한 결과, 10개 품종에서 총 22,450,538개의 변이가 확보되었다. 확보된 변이를 대상으로 한 Filtering을 통해 Machine learning classifier를 이용하여 품종구분 모델을 구축에 활용할 고품질 homo type SNP 총 11,640개 확보하였다. 확보된 11,640개의 SNP 정보를 이용하여 J48 classifier를 이용하여 10-fold cross validation 기준 100% 정확도를 보이는 decision tree를 구축하고 이를 위한 9개의 SNP 마커를 확보하였다. 선발된 SNP780과 SNP867이 13번 염색체에 존재하였으며, SNP 1169와 SNP1271이 10번 염색체에 존재하였으며, 그 밖에 SNP들은 각기 다른 염색체 상에 존재하였다(Table 3). 각 SNP를 이용하여 10개의 품종을 구분하기위하여 decision tree를 작성한 결과(Fig. 1), 9개의 SNP 대한 순차적인 분석을 통해 품종이 구분되는 것으로 확인 되었다.

골든실크와 대황잠 품종 판별

각 품종을 판별하기 위해 선발된 SNP 품종간 차이를 구분하기 위하여 decision tree (Fig. 1)의 가장 상위에 존재하는 SNP부터 순차적으로 검증을 실시하였다. T-ARMS PCR을 통해 SNP780 영역을 증폭하고 전기영동을 통해 밴드의 크기를 분석한 결과(Fig. 2A), 대황잠과 골든실크에서는 175와 265 bp 크기의 밴드가 생성되는 반면, 대박잠을 포함하는 8개 품종에서는 135와 265 bp 크기에 밴드가 형성되어 품종을 두개의 그룹으로 분리할 수 있었다. 이후 대황잠과 골든실크를 세부적으로 구분하기 위하여 SNP867에 대하여 T-ARMS PCR을 수행한 결과(Fig. 2B), 대황잠에서는 501과 376 bp 크기의 밴드가 확인되었으며 골든실크에서는 501과 167 bp 크기가 증폭되어 두개의 품종이 명확하게 구분되었다. 증폭된 SNP 영역에 대한 대립유전자 서열정보를 확인하기 위하여 SNP780 영역에서는 265 bp와 SNP867 영역에서 501 bp 증폭 산물의 염기서열을 분석한 결과(Fig. 2), SNP780에서는 A>T 변이가, SNP867 영역에서는 G>A 변이가 확인되어 두개의 SNP를 이용했을 때 10개 장려품종 중 골든실크와 대황잠을 정확히 구분할 수 있었다.

Table 3. SNP positions and alleles in the whole genome f classification of 10 varieties

SNP	Chromosome	Position	Alleles : varieties
SNP780	Chr13	8628494	T: DaeHwang and GoldenSilk/ A: 8 other varieties
SNP867	Chr13	12322318	A: DaeHwang / G: GoldenSilk
SNP1169	Chr10	7138387	C: DaeBak / T: 7 other varieties
SNP1271	Chr10	11537002	C: KumOk and BaekHwang / T: 5 other varieties
SNP427	Chr4	13708307	T: KumOk / C: BaekHwang
SNP4165	Chr24	9554390	T: JooHwang / C: 4 other varieties
SNP4381	Chr14	3343818	A: HanSaeng / G: 3 other varieties
SNP4545	Chr5	376065	G: BaekOk / T: KumKang and DaeBaek
SNP9183	Chr27	10370819	T: KumKang / A: DaeBaek

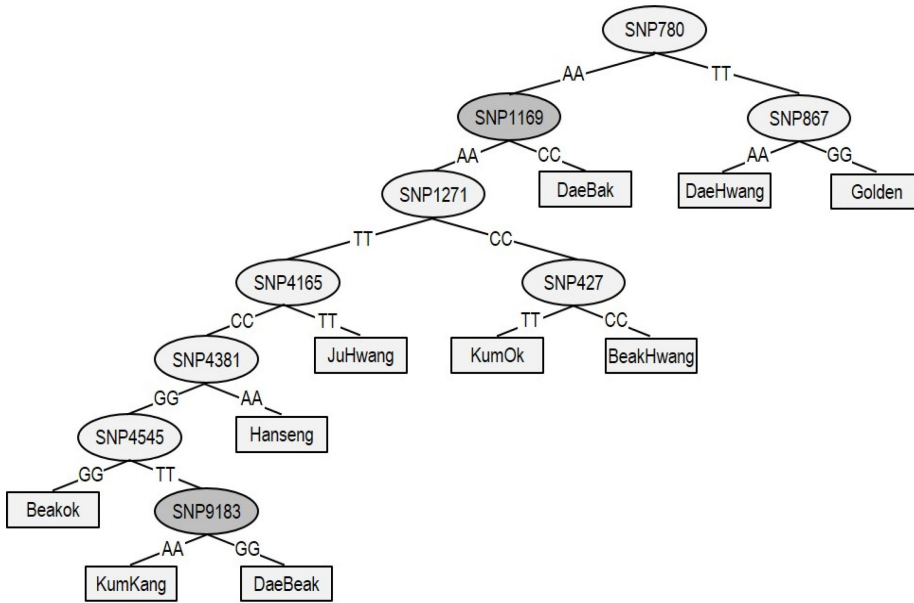


Fig. 1. Whole genome SNP based decision tree for the identification of 10 silkworm varieties. The letter in the circle is the SNP number, the letter in the box is varieties, and the letter in the diagonal line is the allele.

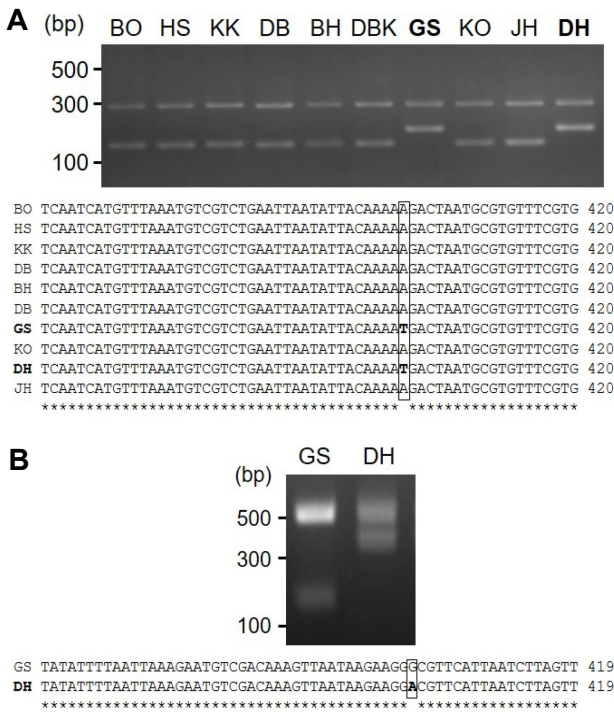


Fig. 2. GoldenSilk and DaeHwang specific SNPs verification using tetra-primer ARMS PCR and sequence alignments of the SNP regions. (A) SNP780 analysis to isolate GoldenSilk and DaeHwang into different groups. (B) SNP867 analysis to differentiate between Golden Silk and DaeHwang. The boxes within the sequence alignments are alleles of each SNPs. BO, BaekOk; HS, HanSaeng; KK, KumKang; DB, DaeBak; BH, Baek Hwang; DBK, DaeBaek; GS, GoldenSilk; KO, Kum Ok.; JH, JooHwang; DH, DaeHwang. Genotyping was performed on 10 individuals per varieties.

PCR-RFLP를 이용한 대박잠 품종 판별

대박잠 품종을 판별하기 위해서, decision tree (Fig. 1) 상에 존재하는 SNP780의 하위에 있는 SNP1169 영역에 대한 증폭 및 제한효소 처리를 통해 품종간 차이를 확인하고 Fig. 3에 나타내었다. 그 결과, SNP1169 영역에 대한 증폭 산물의 크기는 980 bp 이지만, 내부에 제한효소 *Ahd* I의 절단부위가 2번 존재하는 대박잠의 경우 108, 634, 및 238 bp 크기의 세개 밴드가 형성되고, 증폭산물내에 *Ahd*

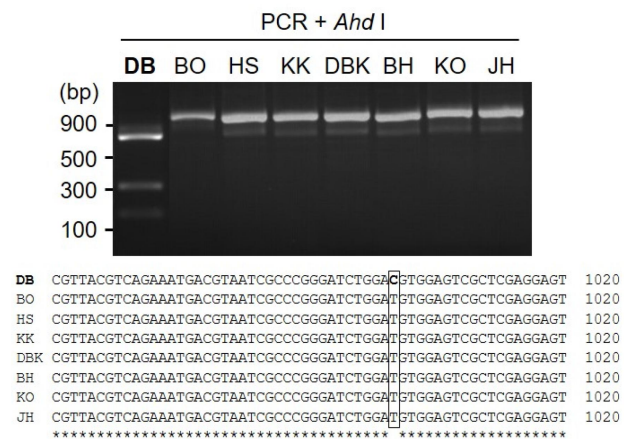


Fig. 3. DaeBak specific SNPs verification using PCR-RFLP and sequence alignments of the SNP regions. To distinguish the DaeBak, SNP1169 was amplified by PCR and compared by treatment with restriction enzyme *Ahd* I. The box within the sequence alignments are alleles of each SNPs. DB, DaeBak; BO, BaekOk; HS, HanSaeng; KK, KumKang; DBK, DaeBaek; BH, BaekHwang; KO, KumOk.; JH, JooHwang. Genotyping was performed on 10 individuals per varieties.

1의 절단부위가 한번만 존재하는 금속잡 포함 7개 품종에서는 108과 872 bp 크기의 밴드 두개만 나타나 대박잡과 차이를 나타냈다(Fig. 3A). 증폭된 SNP1169 영역에 대한 대립유전자를 검증하기 위하여 제한효소를 처리하기 전인 980 bp의 증폭 산물에 대한 염기서열을 분석한 결과, 대박잡에서만 A>C변이가 확인되어 SNP1169 영역에 대한 증폭 및 제한효소 처리를 통해 SNP780 검증에서 분리된 8개 품종 중 대박잡을 정확하게 구분할 수 있었다.

금속잡과 백황잡 품종 판별

Decision tree (Fig. 1) 상의 SNP1169에 의해 하위영역으로 분리된 7개의 품종 중 금속잡과 백황잡을 판별하기 위하여 SNP1271 영역에 대해 T-ARMS PCR을 실시하고 Fig. 4A에 나타내었다. 그 결과, 금속잡과 백황잡에서는 414 및 180 bp의 밴드가 나타난 반면, 주황잡을 포함하는 5개의 품종에서는 414와 278 bp의 밴드가 나타나 밴드 크기를 통해 2개의 그룹으로 분리가 가능하였다. 분리된 그룹 중 금속잡과 백황잡을 각각 판별하기 위하여 SNP427에 대하여 분석한 결과(Fig. 4B), 금속잡은 319와 146 bp의 밴드를 나타내었고, 백황잡은 319와 220 bp의 밴드를 나타내 두개의 품종을 명확히 구분할 수 있었다. 증폭된 SNP

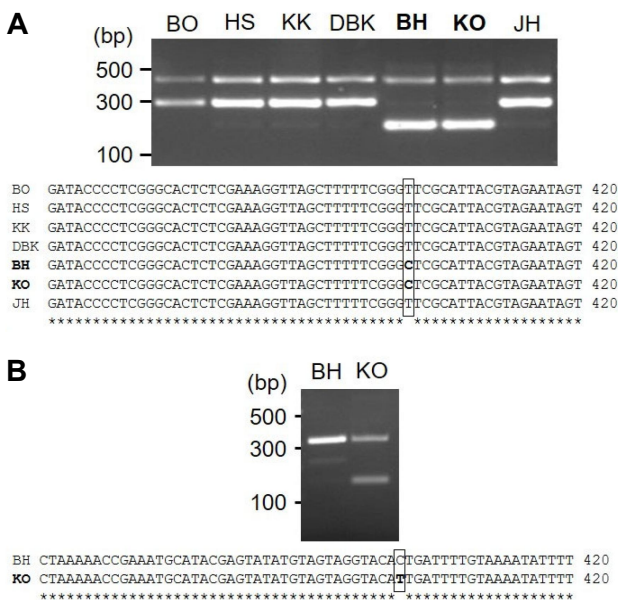


Fig. 4. BaekHwang and KumOk specific SNPs verification using tetra-primer ARMS PCR and sequence alignments of the SNP regions. (A) SNP1271 analysis to isolate BaekHwang and KumOk into different groups. (B) SNP427 analysis to differentiate between BaekHwang and KumOk. The boxes within the sequence alignments are alleles of each SNPs. BO, BaekOk; HS, Han Saeng; KK, KumKang; DBK, DaeBaek; BH, Baek Hwang; KO, KumOk.; JH, JooHwang. Genotyping was performed on 10 individuals per varieties.

영역에 대한 대립유전자 서열정보를 확인하기 위하여 SNP1271영역에서는 414 bp와 SNP427 영역에서 319 bp 증폭 산물의 염기 서열을 분석한 결과(Fig 4), SNP1271에서는 T>C 변이가, SNP427 영역에서는 T>C 변이가 확인되어 두개의 SNP를 이용했을 때, SNP1169에 의해 하위영역으로 분리된 7개의 품종 중 금속잡과 백황잡을 정확히 구분할 수 있었다.

주황잡 및 한생잡 품종 판별

Decision tree (Fig. 1) 상 SNP1271의 하위영역으로 구분되어진 5개의 품종 중 주황잡을 판별하기 위하여 SNP 4165 영역에 대해 T-ARMS PCR을 실시하고 Fig. 5A에 나타내었다. 그 결과, 주황잡에서는 487과 138 bp의 밴드가 나타난 반면 한생잡을 포함하는 4개 품종에서는 487과 392 bp의 밴드가 나타나 주황잡을 구분할 수 있었으며, 487 bp의 증폭 산물에 대한 염기서열 분석 결과 주황잡에서만 C>T의 변이를 가지고 있었다.

주황잡 이외의 4개 품종 중 한생잡을 판별하기 위해서는 SNP4381 영역에 대해 T-ARMS PCR을 실시하여 구분하였다(Fig. 5B). 그 결과, 한생잡에서는 391 및 271 bp의 밴드가 형성되었고 백옥잡을 포함하는 3개 품종에서는

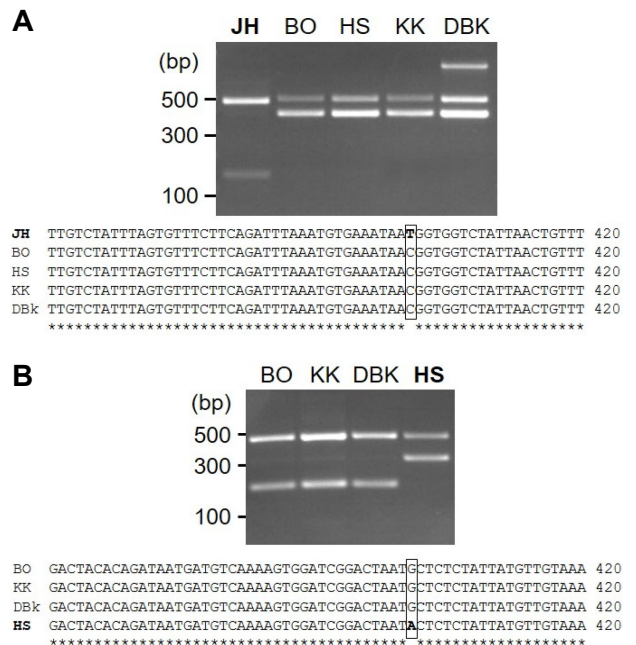


Fig. 5. JooHwang and HanSaeng specific SNPs verification using tetra-primer ARMS PCR and sequence alignments of the SNP regions. (A) SNP4165 analysis to isolate JooHwang into different groups. (B) SNP4381 analysis to isolate HanSaeng into different groups. The boxes within the sequence alignments are alleles of each SNPs. JH, JooHwang; BO, BaekOk; HS, HanSaeng; KK, KumKang; DBK, DaeBaek. Genotyping was performed on 10 individuals per varieties.

391과 164 bp의 밴드가 확인되어 한생잡만을 구분할 수 있었으며, 대립유전자를 검증하기 위하여 391 bp 증폭 산물에 대해 염기서열을 분석한 결과 한생잡에서만 G>T의 변이가 확인 되었다.

백옥잠, 금강잠 및 대백잠 품종 판별

Decision tree (Fig. 1)의 최하위 부분에 존재하는 3개 품종에 대해 판별하기위하여, 백옥잠의 경우 SNP4545 영역을 T-ARMS PCR을 통하여 증폭하였다(Fig. 6A). 그 결과, 백옥잠은 262와 106 bp의 밴드가 증폭되었고 금강잠 및 대백잠에서는 262와 197bp의 밴드가 증폭되어 3개 품종 중 백옥잠만을 구분할 수 있었으며, 대립유전자 검증을 위하여 262 bp의 증폭 산물에 대해 염기서열을 분석한 결과 T>G 변이가 확인되었다. 마지막으로, 금강잠과 대백잠을 구분하기 위해서는 SNP9183 영역에 대한 증폭 후 Ssp I 제한효소를 처리하여 각 품종을 판별하였다(Fig. 6B). SNP9183 영역에 대한 증폭 산물의 크기는 543 bp이며 금강잠의 경우 제한효소부위가 1개만 존재해 109와 434 bp 두개의 밴드가 형성되는 반면 대백잠에서는 2개의 제한효소 부위가 존재해 109, 268, 및 167 bp 세개의 밴드가 나타나 금강잠과 명확하게 구분되었다. 제한효소를 처리하지않은 543 bp의 증폭산물에 대해 염기서열을 분석한

결과, 두 품종에서는 SNP9183 영역에 대립 형질로 A>G를 가지고 있음이 확인되었다.

최근 SNP를 이용하여 농산물의 품종을 판별하고 육종 소재로 이용하기위한 연구가 크게 증가하고 있으나[3, 5, 17, 25], 누에의 경우 다양한 품종을 사육하는 국가가 많지 않아 SNP를 이용한 계통분류를 위한 연구만이 이루어져 있었을 뿐으로[11, 21], 품종판별을 위한 연구가 부족한 실정이었다. 하지만 최근 국내에서 육성 보급을 위한 36 계통의 누에의 전장유전체 서열을 밝히고 표현형을 비교하는 등 유전체 연구를 위한 기반을 갖추고 있다[16]. 담배가루이 및 꿀벌 등에서는 이미 미토콘드리아 SNP를 이용한 구분 및 진단 등이 연구되어 있으나[3, 14], 누에의 경우 보급 장려품종이 1대 교잡종으로 오랜 기간 근친교배로 인해 품종 내에 유전적 유사성을 가지고 있어 단일 SNP 만으로는 정확한 품종 판별이 쉽지 않다[10, 16]. 그뿐만 아니라 모계 유전을 따르는 미토콘드리아 SNP만으로 품종을 판별하기에는 동일한 모체를 가지고 육성된 품종을 선별하는데 제약이 따르게 된다[17]. 하지만 본 연구에서는 각 품종이 가지는 특이적인 SNP를 선별하고, decision tree를 구성하여 2개 이상의 SNP를 이용하여 품종판별의 정확도를 높였다. 이와 같은 방법은 품종판별과정이 다소 복잡해지지만 유전적 유사성을 가진 생물체의 품종판별에서 오류 발생을 최소화하여 품종 판별의 정확도를 높일 수 있을 것으로 판단된다.

감사의 글

본 연구는 2022년도 농촌진흥청 연구사업(세부과제번호: PJ01510703) 및 국립농업과학원 전문연구원 지원사업에 의해 이루어진 것임.

The Conflict of Interest Statement

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Ahn, H. Y., Cha, J. Y., Park, K. R., Kim, Y. R. and Cho, Y. S. 2013. Improvement effect of fermented silkworm (*Bombyx mori* L.) power against orotic acid-induced fatty liver in rats. *J. Life Sci.* **23**, 789-795.
2. Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.
3. Bosco, D., Loria, A., Sartor, C. and Cenis, J. L. 2006. PCR-RFLP identification of *Bemisia tabaci* biotypes in the Mediterranean Basin. *Phytoparasitica* **34**, 243-251.
4. Danecek, P., Auton, A., Abecasis, G., Albers, C. A.,

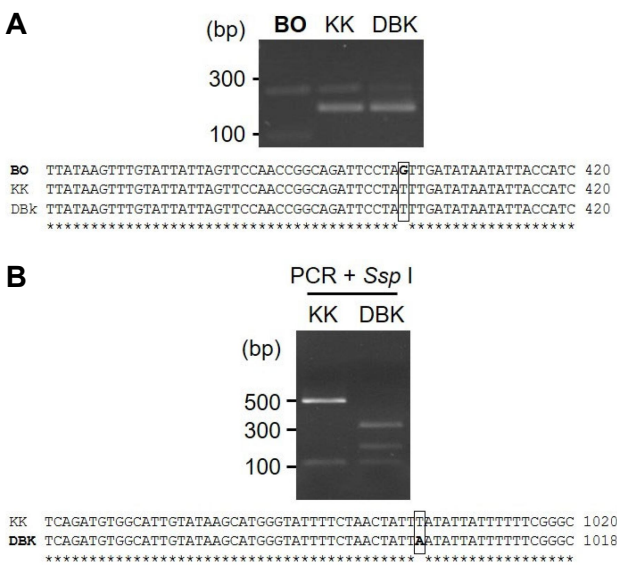


Fig. 6. BaekOk and DaeBaek specific SNPs verification using PCR and sequence alignments of the SNP regions. (A) To isolate the BaekOk, SNP4545 was amplified by using tetra-primer ARMS PCR. (B) To distinguish the DaeBaek, SNP9183 was amplified by PCR and compared by treatment with restriction enzyme Ssp I. The box within the sequence alignments are alleles of each SNPs. BO, BaekOk; KK, KumKang; DBK, DaeBaek. Genotyping was performed on 10 individuals per varieties.

- Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G. and Durbin, R. 2011. The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
5. Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., Negrini, R., Finlay, E. K., Jianlin, H., Groeneveld, E., Weigend, S. and Globaldiv, C. 2010. Genetic diversity in farm animals-a review. *Anim. Genet.* **41**, 6-31.
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. 2009. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* **11**, 10-18.
7. Harada, C. 1961. Heterosis of the quantitative characters in the silkworm. *Bull. Seric. Exp. Sta.* **17**, 1-52.
8. Hong, K. W., Ryu, K. S., Hwang, S. J., Sohn, B. H., Kang, P. D. and Choi, S. R., et al. 1996. Breeding of Kumokjam, an artificial diet adaptable silkworm variety for spring rearing season. *RDA J. Agric. Sci.* **38**, 801-805.
9. Ji, S. D., Kim, N. S., Kweon, H. Y., Choi, B. H., Kim, K. Y. and Koh, Y. H. 2016. Nutrition composition differences among steamed and freeze-dried mature silkworm larval powders made from 3 *Bombyx mori* varieties weaving different colored cocoons. *Int. J. Indust. Entomol.* **33**, 6-14.
10. Kang, P. D., Kim, K. M., Sohn, B. H., Woo, S. O. and Ryu, K. S. 2003. Breeding of a new silkworm variety, Chugangjam, with a sex-limited larval marking and high silk yielding for summer-autumn rearing season. *Int. J. Indust. Entomol.* **6**, 57-61.
11. Kang, P. D., Lee, S. U., Jung, I. H., Shon, B. H., Kim, Y. S., Kim, K. Y., Kim, M. J., Hong, I. P., Lee, K. G. and Park, K. Y. 2007. Breeding of new silkworm variety golden silk, a yellow cocoon color for spring rearing season. *Kor. J. Seric. Sci.* **49**, 14-17.
12. Kang, P. D., Jung, I. Y., Kim, K. Y., Kim, M. J., Sohn, B. H. and Lee, K. G. 2010. Breeding of new silkworm variety with peculiar laval mark "Eolrukmal, Hukpyobeom". *Int. J. Indust. Entomol.* **20**, 115-116.
13. Kim, S. W., Kim, K. Y., Kim, S. R., Kim, S. B., Ji, S. D., Kim, N. S., Kweon, H. Y., Jo, Y. Y. and Kim, J. G. 2018. Breeding of new silkworm variety, 'Chilseongjam' with peculiar larval mark. *Int. J. Indust. Entomol.* **37**, 69-72.
14. Kim, J. S., Kim, M. J., Kim, H. K., Vung, N. N. and Kim, I. 2019. Development of single nucleotide polymorphism markers specific to *Apis mellifera* (Hymenoptera: Apidae) line displaying high hygienic behavior against *Varroa destructor*, an ectoparasitic mite. *J. Asia Pac. Entomol.* **22**, 1031-1039.
15. Kim, M. J., Park, J. S., Kim, H., Kim, S. R., Kim, S. W., Kim, K. Y., Kwak, W. and Kim, I. 2022. Phylogeographic relationships among *Bombyx mandarina* (Lepidoptera: Bombycidae) populations and their relationships to *B. mori* inferred from mitochondrial genomes. *Biology* **11**, 68.
16. Kim, S. W., Kim, M. J., Kim, S. R., Park, J. S., Kim, K. Y., Kim, K. H., Kwak, W. and Kim, I. 2022. Whole-genome sequences of 37 breeding line *Bombyx mori* strains and their phenotypes established since 1960s. *Sci. Data.* **9**, 189.
17. Kowalczyk, M., Staniszewski, A., Kamińska, K., Domaradzki, P. and Horecka, B. 2021. Advantages, possibilities, and limitations of mitochondrial DNA analysis in molecular identification. *Folia Biologica* **69**, 101-111.
18. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
19. Medrano, R. F. V. and Oliveira, C. A. 2014. Guidelines for the tetra-primer ARMS-PCR technique development. *Mol. Biotechnol.* **56**, 599-608.
20. Park, J. W., Lee, C. H., Jeong, C. Y., Kang, S. K., Ju, W. T., Kim, S. W., Kim, N. S., Kweon, H. Y. and Kim, K. Y. 2021. Comparison of antioxidant activity according to silkworm cultivars. *J. Life Sci.* **31**, 1010-1018.
21. Park, J. S., Kim, M. J., Kim, S. W., Kim, K. Y., Kim, S. R. and Kim, I. 2022. Molecular identification of the strains of the domestic silkworm, *Bombyx mori* (Lepidoptera: Bombycidae), which are endemic to Korea, based on single nucleotide polymorphisms in mitochondrial genome sequences. *J. Asia-Pac. Entomol.* **25**, 101922.
22. Partis, L., Croan, D., Guo, Z., Clark, R., Coldham, T. and Murby, J. 2000. Evaluation of a DNA fingerprinting method for determining the species origin of meats. *Meat. Sci.* **54**, 369-376.
23. Ryu, K. S., Lee, H. S., Chung, S. H. and Kang, P. D. 1997. An activity of lowering blood-glucose levels according to preparative conditions of silkworm powder. *Kor. J. Seric. Sci.* **39**, 79-85.
24. Salzberg, S. L. 1994. C4.5: programs for machine learning. *Machine Learning* **16**, 235-240.
25. Son, J. H., Cheong, Y. K., Park, J. C., Kim, Y. K., Kim, K. H., Park, T. I., Kim, B. K. and Kang, C. S. 2017. Construction of complete DNA marker set for 32 Korean wheat cultivar identification. *J. Kor. Int. Agric.* **29**, 150-154.
26. Vasimuddin, M., Misra, S., Li, H. and Aluru, S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium* pp. 314-324.

초록 : 전장유전체 SNP 기반 decision tree를 이용한 누에 품종 판별

박종우^{1*} · 박정선^{2†} · 정찬영¹ · 권혁규¹ · 강상국¹ · 김성완¹ · 김남숙¹ · 김기영¹ · 김익수^{2*}

(¹국립농업과학원 농업생물부, ²전남대학교 응용생물학과)

최근 건강 기능성 식품으로 주목받고 있는 누에가 품종 간 기능성 차이를 나타냄에 따라 품종판별에 대한 필요성이 대두되고 있다. 본 연구에서는 누에의 유전체 내에 존재하는 단일염기다형성(SNP)를 바이오 마커로 이용해 10개의 누에 품종(백황잠, 백옥잠, 대백잠, 대박잠, 대황잠, 골든실크, 향생잠, 주황잠, 금강잠 및 금옥잠)을 판별하기 위하여 전장유전체를 분석하였다. 또한 각 품종 특이적인 SNP를 선발하여 품종을 판별하고자 9개의 SNP를 선발하고 각 품종을 교차 검증 할 수 있는 결정 트리를 작성하여 순차적인 분석을 통한 품종 구분을 실시하였다. 대황잠과 골든실크 그리고 금강잠과 대박잠을 구분하는 각각의 SNP867 및 SNP9183에 대해서는 Restriction fragment length polymorphism을 이용하고, 그 외의 SNP에 대해서는 Tetra-primer Amplification Refractory Mutation System을 이용하여 분석하였다. 그 결과 SNP780부터 SNP9183까지 9개의 SNP를 이용하여 동일 집단을 분리하거나 품종을 선발할 수 있었으며, 해당 영역에 대한 염기서열 분석 결과 대립 유전자가 일치함을 확인하였다. 이러한 결과를 종합해볼 때, 누에 전체 게놈의 SNP와 결정 트리를 이용한 방법은 누에 품종 구분을 위한 판별마커로 이용 가치가 높을 것으로 판단된다.