

<http://dx.doi.org/10.17703/JCCT.2022.8.4.415>

JCCT 2022-7-52 초청논문(Invited Paper)

BCJM 행렬 및 Chargaff 법칙과 Shannon Entropy에 의한 RNA 유전자 비율이 95%이상인 온주감귤과 귤의 유전자 조합

The RNA Base Over 95% of Onju Citrus and Coffee Genes Cut & Paste Based on The BCJM Matrix with Chargaff-Shannon Entropy

이성국*, 김정수**, 이문호***

Sung Kook Lee*, Jeong Su Kim**, Moon Ho Lee***

요약 이기종인 온주감귤 유전자(A=20.57, C=32.71, G=30.01, U=16.71%)와 커피 유전자(A=20.66, C=31.76, G=30.187, U=16.71%)는 95%이상 유전자비가 동일하다. 이기종이면 일반적으로 유전자 결합이 안 되는 것으로 알려졌다. 그러나 유전자 기능적-유사성이 95%이상에서 Chargaff 룰과 Shannon Entropy 조건을 만족하면 접목이 가능하며, 새품종인 Coffrange가 된다. 우리는 DNA-RNA를 세계최초 BCJM 행렬로 풀어 미국특허 및 국제저널에 발표했다. 모든 동식물과 바이러스도 사람이 유전자와 비슷하다. 이점에 착안, 코로나-19와 인체의 유전자 특성을 풀어 영국 행렬교제에 6월 발표했다. 식물에서는 유전자 위치를 쉽게 바꾸는 기법인 BCJM-Transposon으로 처리한다. 시뮬레이션에서는 행렬이 Cut & Paste와 Transpose로 성공할 수 있음을 예측했다.

주요어 : 온주감귤 유전자, 커피 유전자, DNA-RNA

Abstract The heterogeneous Onju citrus genes (A=20.57, C=32.71, G=30.01, U=16.71%) and coffee genes (A=20.66, C=31.76, G=30.187, U=16.71%) have the same genetic ratio of 95% or more. It is known that gene compatibility is generally not possible with this group. However, it can be grafted if the conditions of Chargaff rule and Shannon Entropy are met with gene functional-similarity of more than 95%, and it becomes a new breed of Coffrange. We calculated the world's first BCJM matrix for DNA-RNA and published it in US patents and international journals. All animals and viruses are similar to human genes. Based on this, it was announced in June in the British matrix textbook by solving the genetic characteristics of COVID-19 and the human body. In plants, it is treated with BCJM-Transposon treatment, a technique that easily changes gene location. Simulation predicted that the matrix could be successful with Cut & Paste and Transpose.

Key words : Onju Citrus Gene, Coffee Gene, DNA-RNA

*정회원, 제주국제학교 JEINS 운영지원팀 차장 (제1저자)
(미국 인디애나 주립대 경제학 박사과정 수료)
**정회원, 숭실사이버대학교 ICT공학과 부교수 (참여저자)
***정회원, 전북대학교 전자공학부 초빙교수 (교신저자)
접수일: 2022년 6월 13일, 수정완료일: 2022년 7월 1일
게재확정일: 2022년 7월 8일

Received: June 13, 2022 / Revised: July 1, 2022
Accepted: July 8, 2022
***Corresponding Author: moonho@jbnu.ac.kr
Dept. of Electronics, Chonbuk National Univ. Korea

I. 서론

커피속의 식물은 아프리카와 아시아 열대지방에 약 40 종이 자라지만 흔히 코페아 아라비카(Coffea Arabica)와 코페아 카네포라(Coffea Canephora)를 커피 2대 원종이다. 국제 커피협회(ICO)는 커피를 생산지와 품종에 따라 다음과 같이 분류한다. 코페아 아라비카는 세계 커피 생산량의 70% 이상을 차지하고 원산지는 에티오피아이다. 보통 아라비카 커피(Arabica Coffee)로 부르며 주로 고지대에서 재배되고, 카페인 함량은 1.4% 정도로 낮은 편이다. 재배 조건이 까다로우며 병충해에 취약하지만 맛과 향이 뛰어나다. 코페아 카네포라는 흔히 로부스타(Robusta)라고 부르며 아프리카 콩고가 원산지이다. 아라비카에 비해 저지대에서도 잘 자라고 병충해에 강하지만 맛과 향이 떨어져 주로 블렌딩(Blending) 커피나 인스턴트 커피의 재료로 쓰인다. 카페인 함량은 아라비카보다 2배 정도 높다. 커피나무는 적도를 중심으로 남·북위 25도 사이의 열대 지역에서 생산되며 이 지역을 커피 벨트(Coffee Belt) 또는 커피 존(Coffee Zone)이다. 아라비카(Arabica Coffee)는 까다로운 재배조건을 갖고 있는데 평균 기온은 15~24도 정도가 적합하며 우기와 건기의 구분이 필요하다. 또한 유기질이 풍부하고 배수가 잘되는 화산재 토양, 적당한 햇빛, 800m 이상의 고지대일수록 좋은 제주지역에 알맞다. 반면, 집중호우나 강한 바람은 좋지 않다. 문제는 제주 평균 기온이 17.2도로 커피가 생육하기에는 부적합하다. 기존에 온주밀감은 제주풍토 바람, 습도, 기온에 적응이 되어 노지에 재배되는 토종이 됐다. 제주 노지에 강한 굴과 기후변화에 약한 커피 집목한 Coffrange는 1-3년차까지는 온실에서 배양하고 점차로 노지에 적응, 활성화 된다면 새로운 품종인 Coffrange가 탄생된다. Chargaff Rule과 Shannon Entropy에 의한 RNA 유전자비율이 95%이상인 온주감귤과 귤이 Paste는 Block Circulant Jacket Matrix Code로 Transposon 영역에서 시물레이션하고 노지에서 실험했다 [1-3]. 본 논문의 새로 발견된 식은 (8), (11) 식이 연구의 핵심이 되며, 논문 순서는 1장 서론에 이어, 2장 BCJM 코드, 3장 굴-커피 유전자분석, 4장 Coffrange Transposon 실험, 5장에서 결론을 맺는다.

II. BCJM 코드

1. 정의: $[C]_N = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}$ 는 $N=2p$ 의 2×2 block 행렬이다. 만일 $[C_0]_p$ 와 $[C_1]_p$ 가 $p \times p$ jacket 행렬일때 $[C]_N$ 은 식 (1)의 경우에만 Jacket행렬이 된다 [4-6].

$$C_0 C_1^{RT} + C_1^{RT} C_0 = [0]_N, \quad (1)$$

$[d][C]^{RT} = NI_N$ 인 경우에만 $[C]_N$ 은 Jacket 행렬이다. 그때 C는 다음의 경우에 Jacket 행렬이다. RT(reciprocal transpose)는 역전치이 [7-9].

증명: C_0 와 C_1 이 Jacket 행렬일 때, $C_0 C_0^{RT} = p[I]_p$ 와 $C_1 C_1^{RT} = p[I]_p$ 가 된다. $[C]_N$ 은 $[d][C]^{RT} = NI_N$ 인 경우에만 Jacket Matrix이다. 그때 C는 Jacket 행렬이다.

$$[d][C]^{RT} = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix} \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}^{RT} \\ = \begin{pmatrix} 2p[I]_p & C_0 C_1^{RT} + C_1^{RT} C_0 \\ C_0 C_1^{RT} + C_1^{RT} C_0 & 2p[I]_p \end{pmatrix} = NI_N.$$

그러므로 $[C]_N$ 은 $C_0 C_1^{RT} + C_1^{RT} C_0$ 인 경우에만 Jacket 행렬이다.

예제 2.1: 만일 $C_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, $C_1 = \begin{pmatrix} a & -a \\ -1/a & -1/a \end{pmatrix}$.라 하면 위 식에 의해 $C_0 C_1^{RT} = 2[I]_2$, $C_1 C_1^{RT} = 2[I]_2$, C_0 그리고 C_1 은 Jacket 2차 행렬이다.

그러므로,

$$C_0 C_1^{RT} + C_1^{RT} C_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -1/a & -a \\ -1/a & -a \end{pmatrix} + \begin{pmatrix} a & -a \\ -1/a & -1/a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = [0]_2 \quad (2)$$

$$C_4 = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & a & -a \\ 1 & -1 & -1/a & -1/a \\ a & -a & 1 & 1 \\ -1/a & -1/a & 1 & 1 \end{pmatrix}_{a=1} \\ = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \end{pmatrix} \quad (3)$$

C_0 와 C_1 은 Hadamard 행렬이다. 따라서, 행렬 C_4 는 블록 순환 자켓 행렬이다. C_0 와 C_1 은 아다마르 행렬이다. 순환 서브 행렬의 크기는 2×2 이고, 블록 대각 순환 시프트(block diagonal cyclic shift)의 성질을 가진다. 이러한 서브 행렬을 순환 블록(circulant block)이라고 지칭할 수 있다. 블록 순환 자켓 행렬 C_4 는 다음과 같이 정리될 수 있다.

$$C_4 = I_0 \otimes C'_0 + I_1 \otimes C_1, \quad (4)$$

여기서,
 $I_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $I_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $C'_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, $C_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, \otimes 는 Kronecker 곱이다.

예제 2.2. RNA Upper-lower circulant symmetric 블록 행렬은 [8-10],

$$RNA = \begin{bmatrix} C & U \\ G & A \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}. \quad (5)$$

그러면, Chargaff's rule [8-10]:

$$C+U = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \quad (6)$$

$$A+G = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}. \quad (7)$$

따라서,

$$CU + UC = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} = [0]_2,$$

$$C^2 + U^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 4[I]_2,$$

그리고

$$(C+U) + (\bar{C} + \bar{U}) = (G+A) + (\bar{G} + \bar{A}) = [0]_2,$$

여기서, '—' 는 bit complement이다
 식 6과 7을 결합하면 수정된 Chargaff's rule이 된다.
 즉,

$$(A+G):(C+U) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} ([I]_2 : [I]_2),$$

여기서 하중값은 $\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, normalized하면,

$$(A+G):(C+U) = [I]_2 : [I]_2. \quad (8)$$

(8)식처럼 주어진다.

기존 Chargaff Rule은 (9)와 같다.

$$(A+G):(C+U) = 1:1. \quad (9)$$

기존이 접합이 1:1 직렬(9) 유전자 결합인데, 제안된 BCJM은 유전자가 병렬로 접속되고 '한몸체(I2: I2 단 위행렬)(8) 처리로 정확성과 병렬성으로 신뢰성을 향상시킨다 [1-3].

III. 곱-커피 유전자분석

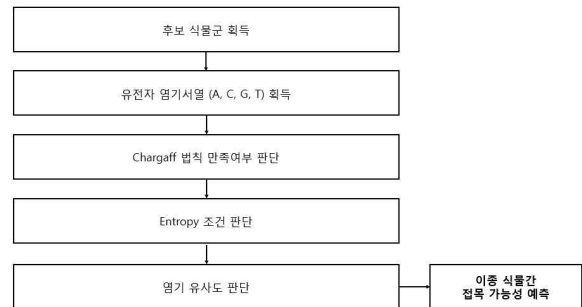


그림 1. 실험 순서도
 Figure 1. Flow chart of experiment

그림1은 일 실험실시에 따른 목(Species)이 다른 식물간 접목 가능성을 예측하는 방법(이하, 예측 방법)을 나타낸다. 예측 방법은 후보 식물군을 획득하는 단계, 후보 식물군의 유전자 염기 서열에 대한 정보를 획득하는 단계, Chargaff 법칙을 만족하는지 여부를 판단하는 단계; 엔트로피 조건을 만족하는지 여부를 판단하는 단계; 및 염기 기능적-유사도를 판단하는 단계를 포함할 수 있다 [1, 2].

다음 표 2를 참조하면, 후보 식물군으로 온주 밀감 나무, 아라비안 커피 나무, 녹차 나무, 탕자 나무, 인삼, 옥수수, 올리브, 녹나무 등이 결정될 수 있다. 결정된 후보 식물군에 대하여 A, C, G, T(U)의 함량의 비율이 도출될 수 있는데, Chargaff 법칙을 만족하는지 여부가 판단될 수 있다.

표 2. 중간 Chargaff 법칙을 따르는 염기 비

Table 2. Ratios of DNA bases conforming to Chargaff's rule

	A	C	G	T	A+G	T+C	
온주밀감	20.57	32.71	30.01	16.71	50.58	49.42	만족
아라비아커피	20.66	31.76	30.87	16.71	51.53	48.47	만족
녹차나무	17.63	34.95	32.83	14.59	50.46	49.54	만족
탱자나무	18.18	34.02	31.96	15.84	50.14	49.86	만족
인삼	23.26	29.55	28.45	18.74	51.71	48.29	만족
옥수수	25.6	24.5	24.6	25.3	50.2	49.8	만족
올리브	22.65	26.87	30.36	20.12	53.01	46.99	
녹나무	18.64	33.9	34.75	12.71	53.39	46.61	

$$H_2(p) = -p \log_2(p) - (1-p) \log_2(1-p) \quad (10)$$

Shannon Entropy 엔트로피는 (10)의 정의에 따라 계산할 수 있다 [1, 7].

표 3. 엔트로피 대칭성 조건 만족 여부

Table 3. Condition check on entropy symmetry

	G+C	entrophy	A+T	entrophy	대칭성
온주밀감	0.6272	0.952798	0.3728	0.952798	만족
아라비아커피	0.6263	0.953471	0.3737	0.953471	만족
녹차나무	0.6778	0.906758	0.3222	0.906758	만족
탱자나무	0.6598	0.92501	0.3402	0.92501	만족
인삼	0.58	0.981454	0.42	0.981454	만족
옥수수	0.491	0.999766	0.509	0.999766	만족

표 3을 참조하면, 엔트로피의 대칭성 조건이 만족되는지 여부가 판단될 수 있다. 예를 들어, 엔트로피의 대칭성 조건은 G+C의 엔트로피와 A+T의 엔트로피가 대칭성을 가지는지 여부일 수 있다.

표 4를 참조하면, 식물의 염기 조성 RNA 확률 변수 $X \sim (0.5, 0.01)$ 일 때, X가 구간 $[a \leq X \leq \beta]$ 또는 $[\beta \leq X \leq a]$ 에 있을 확률을 나타내는 CDF가 계산될 수 있다. 여기서, CDF는 축적 분산 함수(cumulative distribution function)이다.

표 4. 종에 따른 퓨린과 피리미딘 염기의 누적분포함수

Table 4. CDF for DNA bases among plant species

	G+C	A+T	entrophy	CDF
온주밀감	0.6272	0.3728	0.952798	0.7698
아라비아커피	0.6263	0.3737	0.953471	0.7698
녹차나무	0.6778	0.3222	0.906758	0.9281
탱자나무	0.6598	0.3402	0.92501	0.8904
인삼	0.58	0.42	0.981454	0.5762
옥수수	0.491	0.509	0.999766	0
올리브	0.5723	0.4277	0.984864	0.5762
녹나무	0.6865	0.3135	0.897172	0.9281

표 5를 참조하면, CDF 유사도가 판단될 수 있다.

표 5. 중간 퓨린과 피리미딘 염기의 누적분포함수 유사도

Table 5. CDF similarity for DNA bases among plant species

CDF 유사도	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	1	0.829436	0.864555
아라비아커피	1	1	0.829436	0.864555
녹차나무	0.829436	0.829436	1	0.959379
탱자나무	0.864555	0.864555	0.959379	1
인삼	0.748506	0.748506	0.620838	0.647125
옥수수	0	0	0	0
올리브	0.748506	0.748506	0.620838	0.647125
녹나무	0.829436	0.829436	1	0.959379

표 6 ~ 표 9와 같이, 염기 기능유사도를 판단할 수 있다.

표 6. 중간 A 염기의 유사도

Table 6. Adenine similarity among plant species

A	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	0.995644	0.857073	0.883811
아라비아커피	0.995644	1	0.85334	0.879961
녹차나무	0.857073	0.85334	1	0.969747
탱자나무	0.883811	0.879961	0.969747	1
인삼	0.884351	0.88822	0.757954	0.781599
옥수수	0.803516	0.807031	0.688672	0.710156
올리브	0.908168	0.912141	0.778366	0.802649
녹나무	0.906174	0.902227	0.945815	0.975322

표 7. 중간 C 염기의 유사도

Table 7. Cytosine similarity among plant species

C	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	0.970957	0.935908	0.961493
아라비아커피	0.970957	1	0.908727	0.933568
녹차나무	0.935908	0.908727	1	0.973391
탱자나무	0.961493	0.933568	0.973391	1
인삼	0.903393	0.930416	0.845494	0.868607
옥수수	0.749006	0.771411	0.701001	0.720165
올리브	0.821461	0.846033	0.768813	0.78983
녹나무	0.964897	0.936873	0.969957	0.996473

표 8. 중간 G 염기의 유사도

Table 8. Guanine similarity among plant species

G	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	0.972141	0.914103	0.938986
아라비아커피	0.972141	1	0.940299	0.965895
녹차나무	0.914103	0.940299	1	0.9735
탱자나무	0.938986	0.965895	0.9735	1
인삼	0.948017	0.921607	0.866585	0.890175
옥수수	0.819727	0.79689	0.749315	0.769712
올리브	0.988472	0.983479	0.924764	0.949937
녹나무	0.863597	0.888345	0.944748	0.919712

표 9. 종간 T 염기의 유사도

Table 9. Thymine similarity among plant species

T	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	1	0.87313	0.947935
아라비아커피	1	1	0.87313	0.947935
녹차나무	0.87313	0.87313	1	0.921086
탱자나무	0.947935	0.947935	0.921086	1
인삼	0.891676	0.891676	0.778549	0.845251
옥수수	0.660474	0.660474	0.57668	0.626087
올리브	0.830517	0.830517	0.725149	0.787276
녹나무	0.760622	0.760622	0.871145	0.802399

표 10과 같이, 이종 식물간 접목 가능성을 예측할 수 있다.

표 10. 종간 접목 가능성

Table 10. Possibility of grafting different species

T	온주밀감	아라비아커피	녹차나무	탱자나무
온주밀감	1	1	0.87313	0.947935
아라비아커피	1	1	0.87313	0.947935
녹차나무	0.87313	0.87313	1	0.921086
탱자나무	0.947935	0.947935	0.921086	1
인삼	0.891676	0.891676	0.778549	0.845251
옥수수	0.660474	0.660474	0.57668	0.626087
올리브	0.830517	0.830517	0.725149	0.787276
녹나무	0.760622	0.760622	0.871145	0.802399

예시에서 설명의 편의상 접목 가능성이라는 표현을 사용하였으나, 실시 예들은 목이 다른 식물간 유전자 재조합 가능성을 예측하는 데에도 실질적으로 동일하게 적용될 수 있다.

IV. Coffrange Transposon 실험

Transposon은 위치를 쉽게 바꾸는 유전자, 바버라 맥클린톡(1983년 노벨생리학상)이 제안한 옥수수 색깔 전이에서 착상되었다. 커피와 귤이 Cut & Paste 는 유전자 CGUA 상호간 조합을 Transposon 시켜 중의 진화를 이룰 수 있다 [4, 6, 8]. 귤-커피 Transposon 연결은 사람이나 Covid-19의 DNA에서 RNA로 전사되는 과정과 같다 [10]. 그림2에서 9까지 유전자 CUAG도 연결도에 따른 유전자 재조합 식물들을 설명하는 도면들이다. 유전자 재조합이 가능하다고 판단된 후보 식물들(예를 들어, 아라비아 커피 나무와 온주 밀감 나무) 사이에서 유전자 재조합이 이루어질 수 있다.

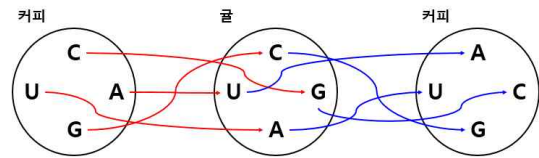
Coffrange Transposon에서 유전자 CGUA 간에 조합 시 Prefix +와 - 표시는 Pyrimidine과 Purine의 조합에

따라 다음 행렬로 구한다.

$$[CGUA]^T \otimes [CGUA] = \begin{bmatrix} CCGGCUCA \\ GCGGUGUA \\ UCUGUUUA \\ ACAGAUAA \end{bmatrix} = \begin{bmatrix} + & + & + & - \\ + & + & + & - \\ + & - & - & - \\ + & - & - & - \end{bmatrix} \quad (11)$$

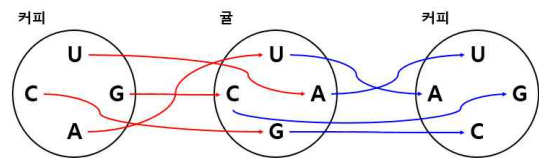
그림 2 ~ 그림 9를 참조하면, 아라비아 커피 내 제1 염기 서열의 시퀀스와 제2 염기 서열의 시퀀스 사이에 온주 밀감 나무와 제3 염기 서열 커피의 시퀀스가 결합될 수 있다.

그림 2와 같이 커피 나무와 귤 나무에 기초한 유전자 재조합 식물은 커피 나무에 포함된 염기 서열 중 CUGA 시퀀스와 AUGC 시퀀스 사이에 귤 나무에 포함된 염기 서열 중 CUAG 시퀀스가 결합된 염기 서열로 결정된다.



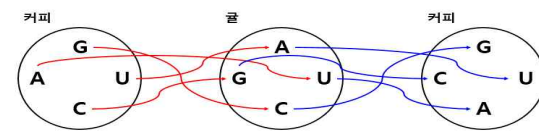
CGC	UAU	GCG	AUA
+	-	+	-

그림 2. 커피, 귤, 커피: CGC-UAU-GCG-AUA
 Figure 2. Coffee, Citrus and Coffee: CGC-UAU-GCG-AUA



UAU	CGC	AUA	GCG
-	+	-	- +

그림 3. 커피, 귤, 커피: UAU-CGC-AUA-GCG
 Figure 3. Coffee, Citrus, and Coffee: UAU-CGC-AUA-GCG



GCG	AUA	CGC	UAU
+	-	+	-

그림 4. 커피, 귤, 커피: GCG-AUA-CGC-UAU
 Figure 4. Coffee, Citrus, and Coffee: GCG-AUA-CGC-UAU

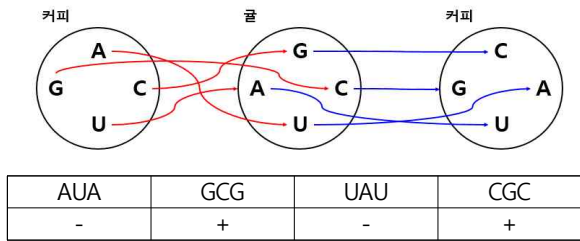


그림 5. 커피,귤,커피: AUA-GCG-UAU-CGC
Figure 5. Coffee, Citrus, and Coffee: AUA-GCG-UAU-CGC

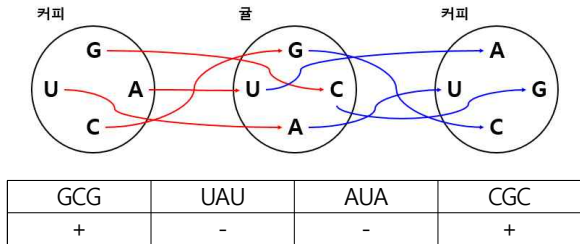


그림 6. 커피,귤,커피: GCG-UAU-AUA-CGC
Figure 6. Coffee, Citrus, and Coffee: GCG-UAU-AUA-CGC

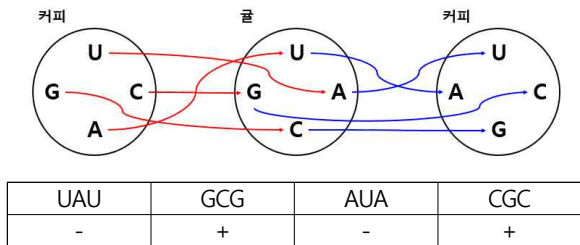


그림 7. 커피,귤,커피: UAU-GCG-AUA-CGC
Figure 7. Coffee, Citrus, and Coffee: UAU-GCG-AUA-CGC

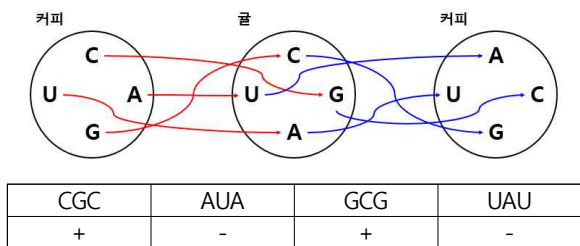


그림 8. 커피,귤,커피: CGC-AUA-GCG-UAU.
Figure 8. Coffee, Citrus, and Coffee: CGC-AUA-GCG-UAU.

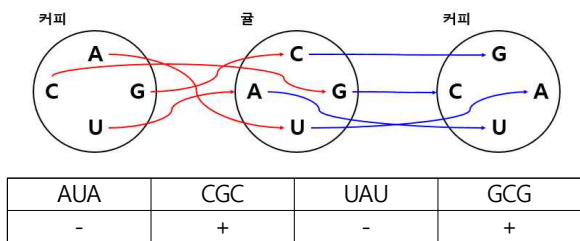


그림 9. 커피,귤,커피: AUA-CGC-UAU-GCG
Figure 9. Coffee, Citrus, and Coffee: AUA-CGC-UAU-GCG

DNA는 같은 축에 두 개의 나선형 사슬을 가지고 있다. 두 사슬은 우회전된 나선을 따르지만, 두 사슬의 연속은 서로 반대 방향으로 진행된다. 본 연구는 생물 정보학 신호 처리 시스템을 통하여 수학적 DNA 이중 나선을 증명한다. 니렌버그 RNA 표준 유전 코드 알파벳의 네 개의 문자들은 64개의 트리플렛(triplet)을 가진다. RNA의 4-문자 알파벳(아데닌 A, 시토신 C, 구아닌 G 및 우라실 U)은 크로네커 계열 $[U C; A G]^{(N)}$ 의 커널로서 $(2*2)^{[U C; A G]}$ 의 형태로 나타낼 수 있는데, 여기서 (N)은 크로네커 멱(power)을 지시한다. 알파벳 $(2*2)$ 행렬의 세 번째 크로네커 멱은 강한 순서로 배치된 64개의 트리플렛의 $(8*8)$ 행렬을 제공한다. 64개의 트리플렛은 단백질의 아미노산을 인코딩한다. $8*8$ Rademacher 행렬은 이중 나선 RNA이며, 각 트리플렛의 처음 두 위치는 트리플렛의 "루트(root)"로 지칭된다. 64개의 트리플렛의 집합은 자연적으로 32개의 트리플렛을 가진 두 개의 동일한 서브셋들로 나뉠 수 있다. (11)식에서 첫 번째 서브셋은 "강한 루트"인 CC, CU, CG, AC, UC, GC, GU, GG를 포함하는 32개의 트리플렛들을 포함한다. "강한 루트"는 +1을 가지는 모든 트리플렛들은 동일한 아미노산을 인코딩하는 것을 의미한다. 두 번째 서브셋은 "약한 루트"인 CC, AA, AU, AG, UA, UU, UG, GA를 포함하는 32개의 트리플렛들을 포함한다. "약한 루트"는 -1을 가지는 모든 트리플렛들은 동일한 아미노산을 인코딩하지 않는 것을 의미한다. 유전자 가위(예를 들어, CRPSPR)로 $8*8$ Rademacher 행렬의 중복 행과 열을 제거한(Cut) 이후, RNA 유전 코드 Transpose해서 R_4 를 출력할 수 있다.

본 연구는 분자 통신의 DNA 또는 RNA 응용에 적용 가능하다. 분자 통신에서 송신기는 정보를 분자 상에 인코딩하고 유체 매체(fluid medium)로 방출한다. 분자는 유체 매체를 통하여 전파되고 수신기는 분자를 수신하여 처리하거나 브라운 운동(Brownian motion)에 의하여 분자와 반응하여 정보를 해독한다.

V. 결론

모든 생물은 DNA 유전자비가 서로 비슷해 사람이 CUAG 20, 30, 30, 20%인데, 커피나 귤은 30, 20, 20, 30%로, 요철특성을 갖는다. 커피나 귤은 30, 20, 20, 30%는 Chargaff 법칙과 Shannon Entropy 법칙에서

99%가 같아 이기종 식물이 유전자 조합에 적당하다. 그 이유는 Forward와 Inverse가 같아 '한몸(Orthogonal Body)'이 되는 것이기 때문이다. BCJM은 이문호가 세계 최초 발견한 행렬로 동식물의 Modular 구조를 만족하는 즉 식물이 잎, 줄기, 가지, 뿌리 등과, 동물이 간, 심장, 위장, 신장처럼 4개 2by2 직교행렬로 구성되어 있는 준직교 블록행렬로 서로 간 간섭을 최소로 하면서 상보성 기능을 최대로 발휘한다. BCJM은 DNA와 RNA와 쉽게 호환 특성과 8by8의 고속알고리즘이 가능하다. 또한, 기준이 접합이 1:1 직렬 유전자 결합인데, 제안된 BCJM은 유전자가 병렬로 접속되고 '한몸체(I2: I2 단위행렬) 처리로 정확성과 병렬성으로 신뢰성을 향상시킨다. 커피-탱자는 엔트로피와 샤가프롤에서 92%로 신뢰성이 떨어졌다. 고려 대상이다. Transposon은 위치를 쉽게 바꾸는 유전자(바버라맥클린 톱 1983년 노벨생리학상)이 제안한 옥수수 색깔 전이에서 착상)인데, 본 연구에서 커피와 꿀이 cut & Paste는 유전자 CUAG 상호간 Transpose 조합을 시켜 종(Species)의 진화를 이룰 수 있다. 염기인자의 +- 구분에 대한 서브셋의 강한뿌리 +와 약한뿌리 -가 왜 그렇게 되는지를 행렬식으로 풀었고 유전자비의 Chargaff Rule을 단위행렬로 풀었다.

1980년대부터 제주감귤은 제주의 주요 농산품, 그런데 기후변화로 한물이 가고 있다. 대치작물은 무엇인가?

Coffrage(꿀+커피)다. 유전자 연구로 가능성을 열어왔다. 생쥐와 인간은 같은 포유동물, 그러나 종이 다르다. 이 두 종(種)이 공통조상에서 갈라져 나온지 오래됐지만 오랜 세월이 지났어도 기능적 유사도(Functional-similarity) 부분은 그다지 분화되지 않아서 서로 비슷하다.

생쥐(A=28, G=22, C=21, T=28%)에서 실험이 성공되면 인체(A=29, G=20, C=20, T=30%)로 곧바로 실행에 옮기듯, 꿀(A=20, C=32, G=30, T=16%)과 커피(A=20, C=31, G=30, T=16%)는 종(Species)이 다르지만 쌍떡잎 식물로 유전체의 기능적 유사도 부분(97%)이 서로 같아 Transposon 영역에서 가능하다.

부 록

(11)식의 Sparse 행렬은 (12)식과 고속 Butterfly는 그림 10과 같고 짝수, 홀수 입력 구분은 그림 11과 같다.

$$\begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix} \\ = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \quad (12)$$

만일, 커피-꿀-커피를 그림 2 ~ 그림 9처럼 $[UCAG]^T \otimes [UCAG] \otimes [UCAG]^T$ Nirenberg Code로

$$\begin{bmatrix} U \\ C \\ A \\ G \end{bmatrix} \otimes [U \ C \ A \ G] \otimes \begin{bmatrix} U \\ C \\ A \\ G \end{bmatrix} = \begin{bmatrix} UUU & UCU & UAU & UGU \\ UUC & UCC & UAC & UGC \\ UUA & UCA & UAA & UGA \\ UUG & UCG & UAG & UGG \\ CUU & CCU & CAU & CGU \\ CUC & CCC & CAC & CGC \\ CUA & CCA & CAA & CGA \\ CUG & CCG & CAG & CGG \\ AUU & ACU & AAU & AGU \\ AUC & ACC & AAC & AGC \\ AUA & ACA & AAA & AGA \\ AUG & ACG & AAG & AGG \\ GUU & GCU & GAU & GGU \\ GUC & GCC & GAC & GGC \\ GUA & GCA & GAA & GGA \\ GUG & GCG & GAG & GGG \end{bmatrix} \begin{bmatrix} -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right) + \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \right) \otimes \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ \Rightarrow \begin{bmatrix} -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \\ -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \end{bmatrix} \\ = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \quad (13)$$

연결하면 (13)식과 같이 되고 decompose 하면 (11), (12), (13)식과 대동소이한 결과를 가져온다. 여기서 Transpose는 유전자 UCAG를 가로를 세로로 움직이는 Transposon (40%) 공간 속에 이뤄진다.

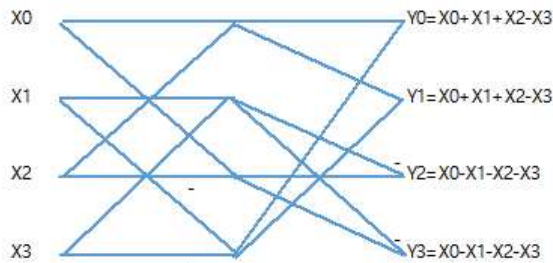


그림 10. 행렬 (12)식의 Butterfly
Figure 10. The Butterfly of Matrix (12)

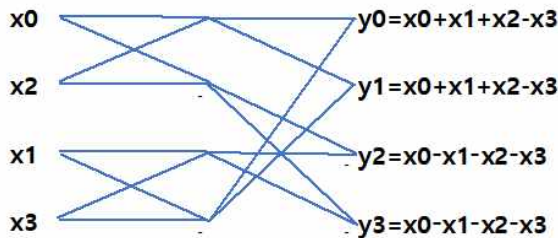


그림 11. 그림 10의 짝, 홀 구분에 의한 Butterfly
Figure 11. Even&Odd Butterfly of Figure 10

References

- [1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, No. 3, pp. 31-423 and pp. 623-656, July 1948, DOI: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- [2] E. Chargaff, S. Zamenhof, and C. Green, "Human desoxyribonucleic acid: Composition of human desoxyribonucleic acid," *Nature*, Vol. 165, pp. 756-757, May 1950, DOI: 10.1038/165756b0.
- [3] J. Watson and F. Crick, "Molecular structure of nucleic acids: A structure for deoxyribonucleic acid," *Nature*, Vol. 171, pp. 737-738, April 1953, DOI: 10.1038/171737a0.
- [4] M.H. Lee and J. Hou, "Fast Block Inverse Jacket Transform," *IEEE Signal Processing Letters*, Vol. 13, No. 8, pp.461-464, August 2006, DOI:10.1109/LSP.2006.873660.
- [5] M.H. Lee, H. Hai and X. Zhang, "MIMO Communication Method and System using the Block Circulant Jacket Matrix," *United States Patent*, US 009356671B1 [Internet]. 31 May 2016. Available from: <https://patentimages.storage.googleapis.com/cb/46/34/4acf23e5a9b6e1/US9356671.pdf> [Accessed: 14 June 2022].
- [6] M.H. Lee, H. Hai, and S.K. Lee, "A Mathematical Design of Nirenberg RNA Standard Genetic Code and Analysis Based on the Block Circulant Jacket Matrix," *Applied to United States Patent*, No. 62/610, 496, 28 December 2017.
- [7] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Process*. 4th ed., Boston, McGraw Hill, 2002.
- [8] M.H. Lee, H. Hai, S.K. Lee, and S. Petoukhov, "A Mathematical Proof of Double Helix DNA to Reverse Transcription RNA for Bioinformatics," *Proceedings of the 1st International Conference of Artificial Intelligence, Medical Engineering, and Education (AIMEE 2017)*, Moscow, Russia, 21 - 23 August 2017, Springer Nature, 2018, DOI: 10.1007/978-3-319-67349-3_5.
- [9] M.H. Lee, S.K. Lee, and K. Cho, "A Life Ecosystem Management With DNA Base Complementarity," *Proceedings of the International Conference of Artificial Intelligence, Medical Engineering, Education (AIMEE 2018)*, Moscow, Russia, 6 - 8 October 2018, Springer Nature, 2020, DOI: 10.1007/978-3-030-12082-5_2.
- [10] S.K. Lee and M.H. Lee, "The COVID-19 DNA-RNA Genetic Code Analysis Using Information Theory of Double Stochastic Matrix," *Matrix Theory - Classics and Advances*, IntechOpen, United Kingdom, 13 June 2022, DOI: 10.5772/intechopen.102342.