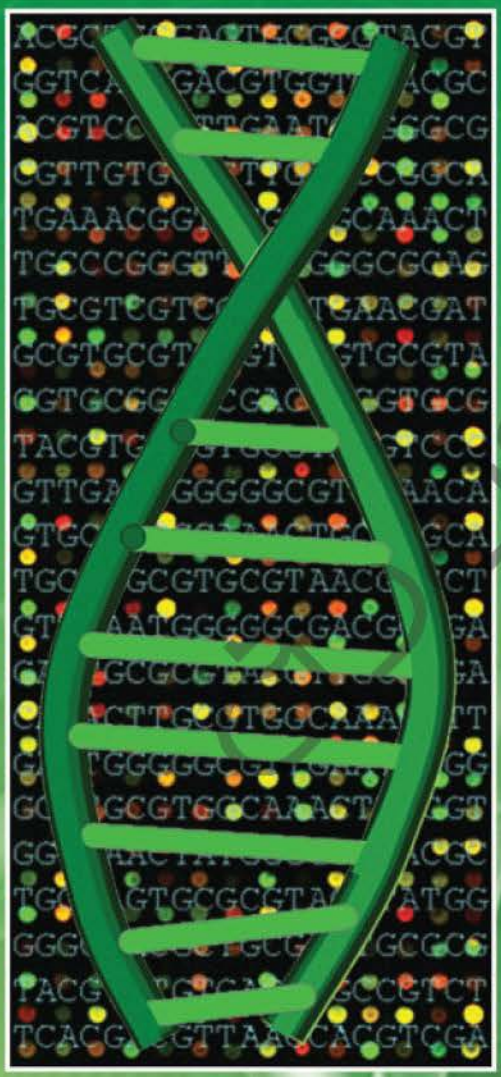
ชีวสารสนเทศทางสัตวศาสตร์



ศุภมิตร เมฆฉาย
ภาควิชาสัตวศาสตร์และสัตว์น้ำ
คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

ประวัติผู้เขียน

ดร. ศุภมิตร เมฆฉาย
ตำแหน่ง รองศาสตราจารย์

หน่วยงานที่ติดต่อ : ภาควิชาสัตวศาสตร์และสัตว์น้ำ
คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่
อำเภอเมือง จังหวัดเชียงใหม่ 50200

การศึกษา

2536 วิทยาศาสตรบัณฑิต (เกษตรศาสตร์) สาขาสัตวบาล คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

2539 วิทยาศาสตรมหาบัณฑิต (เกษตรศาสตร์) สาขาสัตวศาสตร์ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

2546 Dr.agr. (Molecular genetics), University of Bonn, Germany

ประวัติการทำงาน

2540 อาจารย์ประจำภาควิชาสัตวศาสตร์ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

2548 ดำรงตำแหน่งผู้ช่วยศาสตราจารย์ ภาควิชาสัตวศาสตร์ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

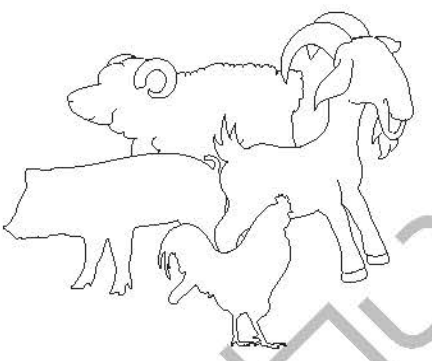
2554 ดำรงตำแหน่งรองศาสตราจารย์ ภาควิชาสัตวศาสตร์และสัตว์น้ำ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

ประสบการณ์ทางวิชาการ

- รับทุน DAAD เพื่อศึกษาระดับปริญญาเอก ที่ Institut für Tierzuchtwissenschaft, University of Bonn, Germany
- รับทุนเลกเปลี่ยนนักวิจัยระหว่างไทย-เยอรมนี เพื่ออบรมเทคโนโลยีชั้นสูงทางด้านจีโนมปศุสัตว์ ที่สถาบัน Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany
- รับทุน EURASIA (Erasmus Mundus, EU) เพื่อแลกเปลี่ยนประสบการณ์การสอนและวิจัยที่ Warsaw University of Life Science (SGGW), Warsaw, Poland
- ดูงานทางด้านการเกษตรและเทคโนโลยีชีวภาพ ที่มหาวิทยาลัย Cornell, New York, USA

ตัวอย่าง

ชีวสารสนเทศทางสัตวศาสตร์



ศุภมิตร เมฆฉาย

ตัวอย่าง

ภาควิชาสัตวศาสตร์และสัตว์น้ำ
คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่
2556

ชีวสารสนเทศทางสัตวศาสตร์

ISBN: 978-974-672-816-4

ผู้แต่ง: ศุภมิตร เมฆฉาย

เจ้าของและผู้จัดพิมพ์: สำนักพิมพ์มหาวิทยาลัยเชียงใหม่
ศูนย์บริหารงานวิจัย มหาวิทยาลัยเชียงใหม่
โทรศัพท์: 0 5394 2671-2, 0 5394 2676
โทรสาร: 0 5394 3600
http://cmupress.cmu.ac.th, E-mail: cmupress.th@gmail.com

พิมพ์ครั้งแรก: ตุลาคม 2556

ราคา: 420 บาท

ข้อมูลทางบรรณานุกรม
ศุภมิตร เมฆฉาย.
ชีวสารสนเทศทางสัตวศาสตร์.-- เชียงใหม่ :
สำนักพิมพ์มหาวิทยาลัยเชียงใหม่, 2556.
272 หน้า.
1. สัตวศาสตร์. 2. ปศุสัตว์. I. ชื่อเรื่อง.
591
ISBN 978-974-672-816-4

ออกแบบปกโดย: พันธภรณ์ สุภัคกาญจน์กุล

ออกแบบและพิมพ์: บริษัท วิทอินดีไซน์ จำกัด

© สงวนลิขสิทธิ์ พ.ศ. 2537 ห้ามลอกเลียนแบบ ทำซ้ำ ดัดแปลง ส่วนหนึ่งส่วนใดของตำราเล่มนี้ รวมทั้งการจัดเก็บถ่ายทอดไม่ว่ารูปแบบหรือวิธีการใดๆ ด้วยกระบวนการทางอิเล็กทรอนิกส์ การถ่ายภาพ การบันทึกหรือวิธีการอื่นใดโดยไม่ได้รับอนุญาต

กรณีต้องการสั่งซื้อเป็นจำนวนมาก กรุณาติดต่อ สำนักพิมพ์มหาวิทยาลัยเชียงใหม่
โทรศัพท์: 0 5394 2671-2, 0 5394 2676 โทรสาร: 0 5394 3600
http://cmupress.cmu.ac.th, E-mail: cmupress.th@gmail.com

คำนำ

หนังสือชีวสารสนเทศทางสัตวศาสตร์ (Bioinformatics in Animal Science) เล่มนี้ ผู้เขียนได้พยายามรวบรวมเอกสาร หนังสือ ตำรา วารสารทางวิชาการ ตลอดจนฐานข้อมูลอิเล็กทรอนิกส์ด้านชีวสารสนเทศที่เกี่ยวข้องในทางสัตวศาสตร์ (ปศุสัตว์) เพื่อเป็นแหล่งข้อมูลให้นักศึกษา และผู้สนใจมีความรู้ ความเข้าใจในชีวสารสนเทศศาสตร์ รหัสพันธุกรรม ฐานข้อมูลทางพันธุกรรม การจัดการข้อมูลทางพันธุกรรม การเปรียบเทียบข้อมูลแบบ Pairwise alignment และ Multiple alignment ฐานข้อมูลจีโนมปศุสัตว์ โดยเฉพาะอย่างยิ่ง ฐานข้อมูล QTL ฐานข้อมูล SNP ฐานข้อมูลจีโนม UCSC และฐานข้อมูลจีโนม Ensembl

ผู้เขียนขอกราบขอบพระคุณครูบาอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาททั้งวิชาและความรู้ โดยเฉพาะรองศาสตราจารย์เพทาย พงษ์เพียจันทร์ ครูผู้จุดประกายสร้างความสนใจทางด้านคอมพิวเตอร์ให้แก่ผู้เขียน ผู้เขียนขอกราบขอบพระคุณ Prof. Dr. Siriluck Ponsuksili และ Prof. Dr. Klaus Wimmers ซึ่งเป็นครูผู้แนะนำให้ผู้เขียนรู้จักศาสตร์ของชีวสารสนเทศเป็นครั้งแรก และขอกราบขอบพระคุณ Prof. Dr. Karl Schellander, University of Bonn, Germany ที่ให้การสนับสนุนทางด้านวิชาการเป็นอย่างดีเยี่ยม นอกจากนี้ใคร่ขอขอบคุณกำลังใจและแรงผลักดันจากครอบครัวที่เป็นกำลังสำคัญ ในการเขียนหนังสือเล่มนี้ให้สำเร็จลุล่วงไปได้ด้วยดี ขอขอบคุณภาควิชาสัตวศาสตร์และสัตว์น้ำ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่สนับสนุนและให้โอกาสที่ดีมาโดยตลอด ขอขอบคุณนักศึกษาปริญญาโท และปริญญาเอก ที่เป็นกำลังสำคัญในการพิสูจน์อักษรของหนังสือเล่มนี้ให้ถูกต้อง สุดท้ายนี้ขอขอบพระคุณผู้ทรงคุณวุฒิภายนอกทั้ง 2 ท่าน รวมถึง รศ. ดร. จีรยุทธ ไชยจารุวณิช ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ให้ข้อแนะนำ และช่วยตรวจสอบแก้ไขต้นฉบับหนังสือเล่มนี้ให้สมบูรณ์อย่างดียิ่ง

ผู้เขียนหวังว่า หนังสือเล่มนี้จะเป็นประโยชน์ต่อการเรียนรู้ของนักศึกษา และผู้สนใจ ผู้เขียนยินดีน้อมรับข้อเสนอแนะ และข้อท้วงติงทั้งหมด เพื่อทำให้เนื้อหาวิชาการของหนังสือเล่มนี้ มีความสมบูรณ์ยิ่งขึ้น

ศุภมิตร เมฆฉาย

ตุลาคม 2556

คำนิยม

หนังสือเรื่อง ชีวสารสนเทศศาสตร์ทางสัตวศาสตร์ ซึ่งเรียบเรียงโดย รศ.ดร.ศุภมิตร เมฆฉาย อาจารย์ประจำภาควิชาสัตวศาสตร์และสัตว์น้ำ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่ ถือเป็นอีกหนึ่งผลงานที่เต็มเปี่ยมด้วยคุณค่าทางวิชาการ และเทคนิคเชิงลึกทางชีวสารสนเทศศาสตร์ ผู้เขียนได้ให้ความรู้ความเข้าใจพื้นฐานเกี่ยวกับรหัสพันธุกรรมของสิ่งมีชีวิตและการใช้ประโยชน์จากข้อมูลรหัสพันธุกรรม โดยเฉพาะอย่างยิ่งด้านสัตวศาสตร์เพื่อการค้นหาตำแหน่งของยีนที่มีส่วนควบคุมหรือมีความเกี่ยวพันกับลักษณะสำคัญทางเศรษฐกิจ เช่น การให้ผลผลิต การเจริญเติบโต ความสมบูรณ์พันธุ์ ลักษณะสุขภาพ หรือ ภูมิต้านทานต่อโรค ของสัตว์เศรษฐกิจ เช่น สุกร ไก่ โค หรือ แกะ เป็นต้น

ผู้เขียนเริ่มต้นโดยการแนะนำให้ผู้อ่านรู้จักฐานข้อมูลสำคัญหลักที่รวบรวมและเผยแพร่ข้อมูลรหัสพันธุกรรมของสิ่งมีชีวิตชนิดต่างๆ รวมถึงรูปแบบการจัดเก็บข้อมูล การสืบค้นและการเปรียบเทียบส่วนเหมือนส่วนต่างของรหัสพันธุกรรมที่ปรากฏในสิ่งมีชีวิตชนิดต่างๆ ด้วยโปรแกรมคอมพิวเตอร์ทางชีวสารสนเทศศาสตร์ที่มีให้บริการโดยทั่วไป โดยแสดงตัวอย่างและอธิบายรายละเอียดทางเทคนิคของการใช้งานอย่างครบถ้วนสมบูรณ์และง่ายต่อการทำความเข้าใจ

หลังจากนั้นผู้เขียนได้จำเพาะให้รู้จักฐานข้อมูลรหัสพันธุกรรมของสัตว์เศรษฐกิจต่างๆ ที่มีการเผยแพร่ โดยเฉพาะอย่างยิ่ง แผนที่ทางพันธุกรรมและเทคนิคการสืบค้นตำแหน่งของเครื่องหมายโมเลกุลดีเอ็นเอที่พบว่ามีความสัมพันธ์กับความผันแปรเชิงปริมาณของลักษณะสำคัญทางเศรษฐกิจ รวมถึงฐานข้อมูลเครื่องหมายโมเลกุล SNP ที่ระบุตำแหน่งรหัสพันธุกรรมที่มีความผันแปรอย่างมีนัยสำคัญระหว่างประชากรของสิ่งมีชีวิตและอาจมีความสัมพันธ์กับความผันแปรเชิงปริมาณของลักษณะสำคัญทางเศรษฐกิจ นอกจากนั้นยังแนะนำฐานข้อมูลอื่นๆ ที่เกี่ยวข้องเพิ่มเติมอีกด้วย

จากภาพรวมดังกล่าวจะเห็นได้ว่าผู้เขียนมีความประสงค์จะสร้างความเข้าใจแก่ผู้อ่านตั้งแต่พื้นฐานและเจาะลึกให้รู้จักการใช้ประโยชน์จากข้อมูลรหัสพันธุกรรมที่มีความเกี่ยวพันกับลักษณะสำคัญทางเศรษฐกิจของสัตว์เศรษฐกิจชนิดต่างๆ ได้อย่างถี่ถ้วน หนังสือเล่มนี้จึงเป็นประโยชน์และเหมาะแก่การเรียนรู้ของนักศึกษา นักวิจัย นักวิทยาศาสตร์ คณาจารย์ ที่ศึกษาวิจัยด้านสัตวพันธุศาสตร์เป็นอย่างยิ่ง

รศ.ดร.จีรยุทธ ไชยจารุวณิช
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

สารบัญ

		หน้า
คำนำ		**i**
คำนิยม		**ii**
สารบัญ		**iii**
คำย่อ		**vi**
บทที่ 1	**ชีวสารสนเทศศาสตร์**	**1**
	1.1 ความหมายของชีวสารสนเทศศาสตร์	1
	1.2 ขอบเขตของชีวสารสนเทศศาสตร์	2
	1.3 การประยุกต์ใช้ชีวสารสนเทศศาสตร์	4
	1.4 ชีวสารสนเทศทางสัตวศาสตร์	6
บทที่ 2	**รหัสพันธุกรรม**	**7**
	2.1 ข้อมูลรหัสพันธุกรรม	7
	2.2 การแปลรหัสพันธุกรรมไปเป็นโปรตีน	12
	2.3 การถอดรหัสพันธุกรรม	13
	2.4 การใช้ประโยชน์จากข้อมูลรหัสพันธุกรรม	16
บทที่ 3	**ฐานข้อมูลรหัสพันธุกรรม**	**19**
	3.1 ฐานข้อมูลลำดับนิวคลีโอไทด์	19
	3.2 การส่งข้อมูลลำดับนิวคลีโอไทด์เข้าสู่ฐานข้อมูล	25
	3.3 การค้นหาข้อมูลลำดับนิวคลีโอไทด์	28
	3.4 ฐานข้อมูลทางชีวโมเลกุลของสิ่งมีชีวิต	41
บทที่ 4	**การจัดการข้อมูลรหัสพันธุกรรม**	**43**
	4.1 การเปลี่ยนรูปแบบการแสดงผลลำดับนิวคลีโอไทด์	43
	4.2 การกลับทิศทางลำดับนิวคลีโอไทด์	48
	4.3 การกลับทิศทางลำดับนิวคลีโอไทด์ของเบสคู่สม	49
	4.4 การจัดเก็บข้อมูลลำดับนิวคลีโอไทด์ในรูปแบบเอกสาร	52
	4.5 การแปลรหัสพันธุกรรมไปเป็นกรดอะมิโน	54

บทที่ 5 การเปรียบเทียบข้อมูลแบบ Pairwise Alignment **65**
5.1 โปรแกรม BLAST 65
5.2 โปรแกรม Basic BLAST 70
5.3 ตัวอย่างการใช้โปรแกรม BLASTN 81
5.4 ตัวอย่างการใช้โปรแกรม BLASTP 89
5.5 ตัวอย่างการใช้โปรแกรม BLAST 2 Sequence 90
5.6 ตัวอย่างการใช้โปรแกรม BLAST Assembled RefSeq Genomes 93

บทที่ 6 การเปรียบเทียบข้อมูลแบบ Multiple Alignment **97**
6.1 Multiple alignment 97
6.2 หลักการของการเปรียบเทียบข้อมูลแบบ Multiple alignment 98
6.3 ตัวอย่างการใช้โปรแกรม Multiple alignment 100

บทที่ 7 ฐานข้อมูลจีโนมของสัตว์เศรษฐกิจ **115**
7.1 จีโนมของสัตว์เศรษฐกิจ 115
7.2 ฐานข้อมูลจีโนมสัตว์เศรษฐกิจ 118

บทที่ 8 ฐานข้อมูล QTL **143**
8.1 การวิเคราะห์ QTL 143
8.2 ฐานข้อมูล QTL ของสัตว์เศรษฐกิจ 144
8.3 การสืบค้นฐานข้อมูล QTLdb 158
8.4 ฐานข้อมูล Bovine QTL Viewer 169

บทที่ 9 ฐานข้อมูล SNP **179**
9.1 เครื่องหมายโมเลกุล SNP 179
9.2 ประเภทของความผันแปรทางพันธุกรรม 181
9.3 ฐานข้อมูล SNP 181
9.4 ตัวอย่างการสืบค้นข้อมูล SNP 184

บทที่ 10 ฐานข้อมูลจีโนม UCSC **195**
10.1 ฐานข้อมูล USCS 195
10.2 องค์ประกอบของฐานข้อมูลจีโนม UCSC 196
10.3 ตัวอย่างการสืบค้นฐานข้อมูลจีโนม UCSC 198

บทที่ 11 ฐานข้อมูลจีโนม Ensembl **209**

11.1 ฐานข้อมูล Ensembl 209

11.2 องค์ประกอบของฐานข้อมูล Ensembl 210

11.3 การสืบค้นข้อมูลจีโนมจากฐานข้อมูล Ensembl 214

11.4 ตัวอย่างการใช้โปรแกรม BLAST/BLAT ในฐานข้อมูล Ensembl 230

11.5 ตัวอย่างการใช้โปรแกรม Biomart ในฐานข้อมูล Ensembl 234

บรรณานุกรม **239**

ภาคผนวก **252**

ดัชนี **254**

อภิธานศัพท์ **260**

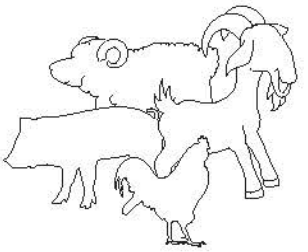
ตัวอย่าง

คำย่อ

A	adenine
AFLP	amplied fragment length polymorphism
BBSRC	Biotechnology and Biological Sciences Research Council
BCM	Baylor College of Medicine
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-Like Alignment Tool
bp	base pair
C	cytosine
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CIB-DDBJ	Center for Information Biology and DNA Data Bank of Japan
cM	centimorgan
COBALT	Constraint-Based Muliple Alignment Tool
dbSNP	database Single Nucleotide Polymorphism
DDBJ	DNA Data Bank of Japan
ddNTPs	dideoxynucleotide triphosphte
DNA	deoxyribonucleic acid
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archive
ENCODE	encyclopedia of DNA elements
EST	expressed sequence tag
G	guanine
GSS	genome survey sequences
HTG	high throughput genome sequence
INSDC	International Nucleotide Sequence Database Collaboration
IUPAC	International Union of Pure and Applied Chemistry
kb	kilobase
MAF	minor allele frequency

Mb	megabase
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
NIG	National Institute of Genetics
NIH	National Institutes of Health
nm	nanometre
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
PIR	Protein Information Resource
QTL	quantitative trait loci
refSNP	reference single nucleotide polymorphism
RH	radiation hybrid
RNA	ribonucleic acid
SAGE	serial analysis of gene expression
SNP	single nucleotide polymorphism
SRA	sequence read archive
STR	short tandem repeat
STS	sequence tagged site
T	thymine
TPA	third party annotation
tRNA	transfer ribonucleic acid
UCSC	University of California Santa Cruz
UTR	untranslated region
WGS	whole genome shotgun

ebeju

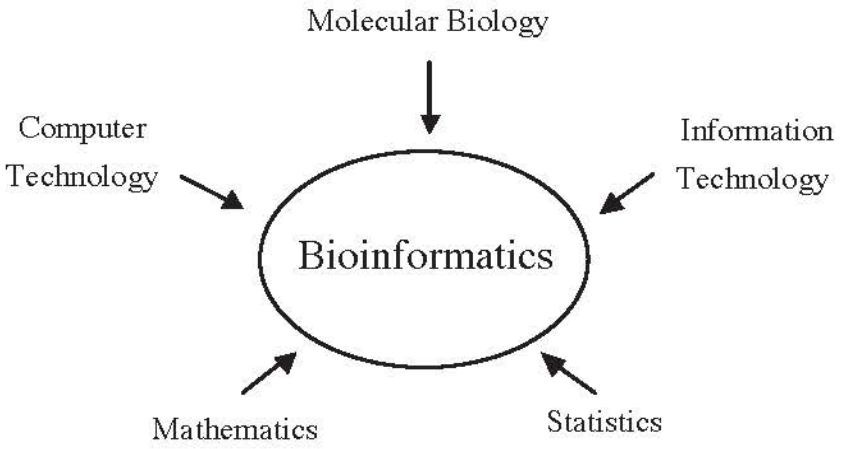


บทที่ 1
ชีวสารสนเทศศาสตร์

ในบทนี้จะกล่าวถึง ความหมายของชีวสารสนเทศศาสตร์ ขอบเขตและการประยุกต์ใช้ชีวสารสนเทศทางด้านต่าง ๆ เช่น วิทยาศาสตร์พื้นฐาน การแพทย์ เภสัชศาสตร์ นิติวิทยาศาสตร์ เทคโนโลยีชีวภาพ และการเกษตร โดยเฉพาะชีวสารสนเทศทางสัตวศาสตร์

1.1 ความหมายของชีวสารสนเทศศาสตร์

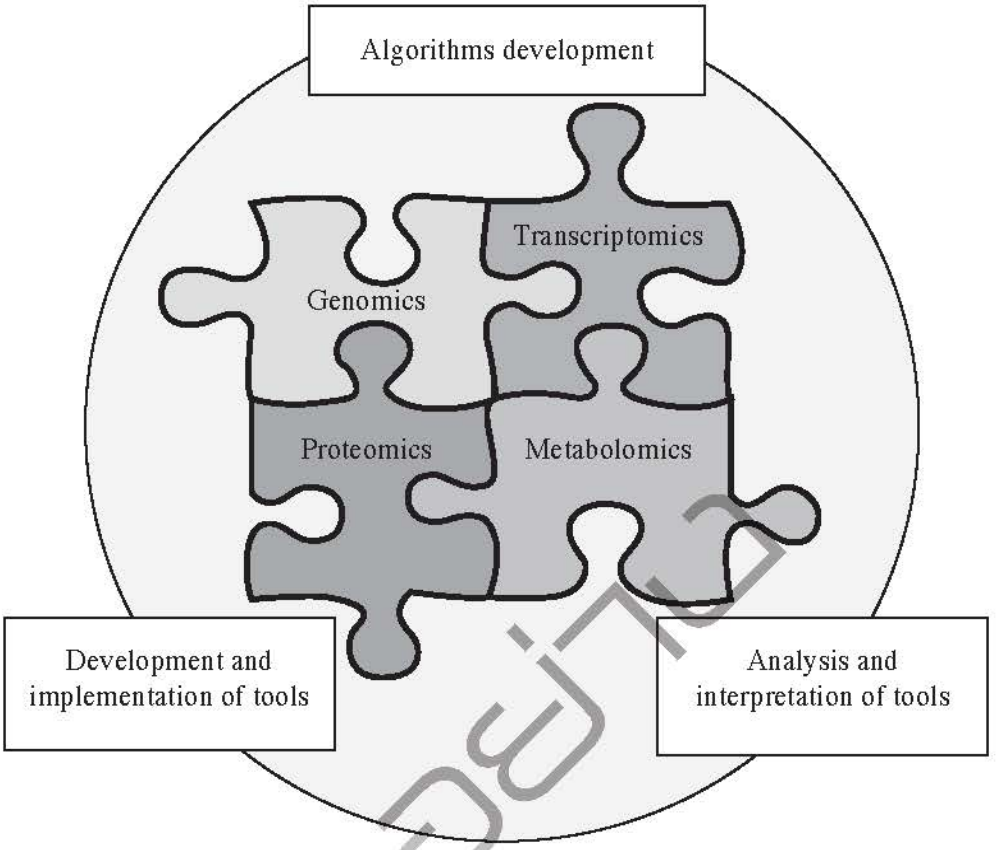
ชีวสารสนเทศ หรือชีวสารสนเทศศาสตร์ (Bioinformatics) เป็นการบูรณาการสหวิทยาการทางด้านคอมพิวเตอร์ เทคโนโลยีทางสารสนเทศ สถิติ และคณิตศาสตร์ เข้ากับองค์ความรู้ทางชีวโมเลกุล (Molecular biology) (ภาพที่ 1-1) เพื่อใช้จัดการและประมวลผลข้อมูลทางชีวโมเลกุลของสิ่งมีชีวิต รวมถึงเก็บรวบรวม วิเคราะห์ สืบค้น เชื่อมโยง และถ่ายโอนข้อมูลทางเครือข่ายคอมพิวเตอร์ (Xiong, 2006; Harisha, 2007; Sharma et al., 2008; Ramsden, 2009) หรือกล่าวอีกนัยหนึ่งได้ว่า ชีวสารสนเทศศาสตร์ คือ การใช้เทคโนโลยีทางคอมพิวเตอร์ (Computational technology) และเทคโนโลยีทางสารสนเทศ (Informatic technology) ร่วมกับสถิติ และคณิตศาสตร์ เพื่อวิเคราะห์ข้อมูลทางชีวโมเลกุล ทั้งในระดับ genome (DNA), transcriptome (RNA), proteome (protein) และ metabolome (metabolites) รวมถึงการทำนายโครงสร้างและหน้าที่ของยีน โปรตีน และสารชีวโมเลกุลของสิ่งมีชีวิต สำหรับการวิเคราะห์ข้อมูลทางชีวโมเลกุลโดยอาศัยเครื่องมือทางชีวสารสนเทศ เรียกอีกชื่อหนึ่งว่า การวิเคราะห์ข้อมูลแบบ *in silico*



ภาพที่ 1-1: การบูรณาการสหวิทยาการของสาขาวิชาที่เกี่ยวข้องกับชีวสารสนเทศศาสตร์

1.2 ขอบเขตของชีวสารสนเทศศาสตร์

ชีวสารสนเทศศาสตร์ในทางชีววิทยาโมเลกุลของสิ่งมีชีวิตนั้น สามารถแบ่งออกเป็น 4 ระดับ คือ (1) จีโนมิกส์ (genomics) (2) ทรานสคริปโตมิกส์ (transcriptomics) (3) โปรตีโอมิกส์ (proteomics) และ (4) เมแทบอโลมิกส์ (metabolomics) (ภาพที่ 1-2) โดยที่จีโนมิกส์ เป็นการศึกษาลำดับจีโนม (genome sequence) หรือโครงสร้างของจีโนม (structural genomics) ในขณะที่ทรานสคริปโตมิกส์ เป็นการศึกษาหน้าที่การทำงานของจีโนม (functional genomics) หรือการแสดงออกของยีน (gene expression) ส่วนโปรตีโอมิกส์ เป็นการศึกษาผลผลิตของยีน ที่มีการแสดงออกมาในเซลล์หรือในเนื้อเยื่อเป้าหมาย สำหรับเมแทบอโลมิกส์ เป็นการศึกษาสารเมแทบอไลท์ (metabolites) ทั้งหมดที่เกิดขึ้นภายในเซลล์ เพื่อศึกษาระบบชีววิทยา (systems biology) ของสิ่งมีชีวิต



ภาพที่ 1-2: ขอบเขตของชีวสารสนเทศศาสตร์

สำหรับขอบเขตของงานชีวสารสนเทศศาสตร์ สามารถแบ่งออกเป็น 3 กลุ่มงาน ซึ่งประกอบด้วย

(1) กลุ่มงานที่ใช้เทคโนโลยีทางคอมพิวเตอร์และสารสนเทศ สำหรับพัฒนาขั้นตอนวิธีการแก้ไขปัญหาแบบใหม่ ๆ โดยใช้ความรู้ทางคณิตศาสตร์ และสถิติ ซึ่งเรียกว่า การพัฒนาอัลกอริทึม (algorithm) รวมถึงการสร้างแบบจำลองทางคอมพิวเตอร์ (computer simulation) เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ ระบุตำแหน่งยีนบนโครโมโซม ทำนายโครงสร้างและหน้าที่ของยีน รวมทั้งการจัดจำแนกชนิดกลุ่มของยีนหรือโปรตีน

(2) กลุ่มงานที่ใช้ความรู้ทางชีวสารสนเทศเป็นเครื่องมือสำหรับศึกษาวิเคราะห์ข้อมูลลำดับนิวคลีโอไทด์ โครงสร้างของจีโนมิกส์ (structural genomics) การทำงานของจีโนมิกส์ (functional genomics) และการเปรียบเทียบพันธุกรรมระหว่างสิ่งมีชีวิต (comparative genomics)

(3) กลุ่มงานที่พัฒนาและใช้เครื่องมือทางชีวสารสนเทศ เพื่อเพิ่มประสิทธิภาพในการใช้และจัดการข้อมูลทางชีววิทยาโมเลกุลที่อยู่ในรูปแบบแตกต่างกัน หรือพัฒนาเครื่องมือใหม่ เพื่อบูรณาการ

ข้อมูลจากแหล่งต่าง ๆ เข้าด้วยกัน ปรากฏเป็นมิติใหม่ของการแสดงผล ซึ่งทำให้เกิดความรู้ความเข้าใจกลไกในระบบชีววิทยาของสิ่งมีชีวิตมากขึ้น

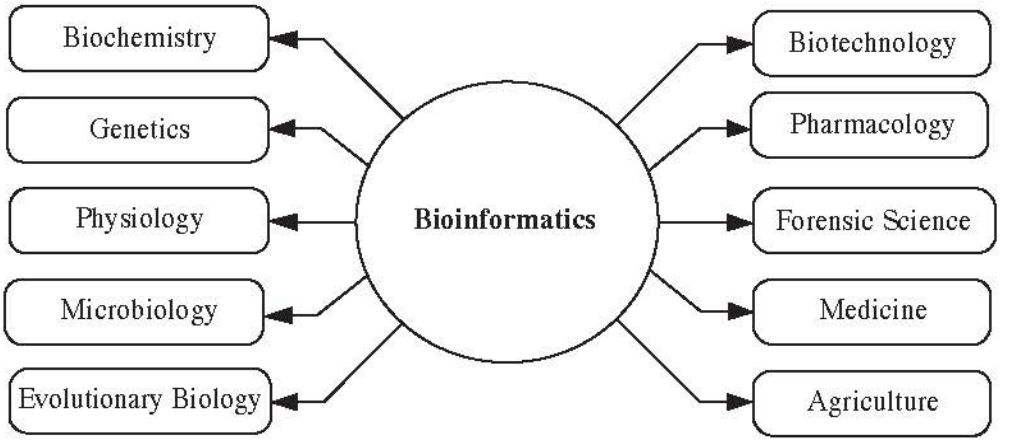
1.3 การประยุกต์ใช้ชีวสารสนเทศศาสตร์

ชีวสารสนเทศศาสตร์ ถูกนำไปประยุกต์ใช้ในศาสตร์สาขาต่าง ๆ มากมาย (ภาพที่ 1-3) ทั้งในด้านวิทยาศาสตร์พื้นฐาน และวิทยาศาสตร์ประยุกต์ ซึ่งแต่ละสาขามีวัตถุประสงค์ที่แตกต่างกัน โดยมีรายละเอียดดังนี้

1.3.1 ทางด้านวิทยาศาสตร์พื้นฐาน ความรู้ทางด้านชีวสารสนเทศศาสตร์ ถูกนำมาใช้ใน ทางวิทยาศาสตร์พื้นฐานในสาขาวิชาต่าง ๆ อาทิเช่น ชีวเคมี พันธุศาสตร์ จุลชีววิทยา สรีรวิทยา และวิวัฒนาการในเชิงชีววิทยา เพื่อสร้างองค์ความรู้พื้นฐานของสิ่งมีชีวิตต่าง ๆ ทั้งพืช สัตว์ และจุลินทรีย์ เพื่อนำไปใช้กับวิทยาศาสตร์เชิงประยุกต์ต่อไป

1.3.2 ทางด้านการแพทย์ ผลจากการศึกษาจีโนมของมนุษย์ ทำให้ทราบข้อมูลลำดับนิวคลีโอไทด์ของมนุษย์ทั้งหมด โดยชีวสารสนเทศศาสตร์ สามารถนำมาช่วยค้นหายีนที่เป็นสาเหตุของโรคผิดปกติทางพันธุกรรมได้เป็นจำนวนมาก รวมถึงโรคที่มีความซับซ้อน (complex disease) ซึ่งเกิดจากปฏิสัมพันธ์ระหว่างสารพันธุกรรมและสิ่งแวดล้อม เช่น โรคมะเร็ง โรคเบาหวาน และโรคหัวใจ โดยชีวสารสนเทศ สามารถช่วยสืบค้น วิเคราะห์และทำนาย ปัจจัยการเปลี่ยนแปลงในระดับเซลล์ ทำให้เข้าใจกลไกทางชีววิทยาโมเลกุลของโรคเหล่านี้มากขึ้น ซึ่งนำไปสู่การค้นพบองค์ความรู้ใหม่ สำหรับพัฒนาแนวทางการป้องกัน และรักษาโรคดังกล่าวได้อย่างมีประสิทธิภาพมากขึ้น นอกจากนี้ความรู้ทางชีวสารสนเทศศาสตร์ ยังถูกนำมาประยุกต์ใช้ในการผลิตวัคซีนเพื่อป้องกันโรคอุบัติใหม่ การรักษาโรคด้วยวิธียีนบำบัด (gene therapy) หรือการตรวจวินิจฉัยโรคในระดับโมเลกุล (molecular diagnosis) เป็นต้น

1.3.3 ด้านเภสัชศาสตร์ ความรู้ทางชีวสารสนเทศศาสตร์ ถูกนำมาใช้ค้นหาโมเลกุลเป้าหมายสำหรับใช้ออกแบบยารักษาโรค ทำให้เกิดการพัฒนายาชนิดใหม่ขึ้นมา โดยอาศัยความรู้พื้นฐานของโปรตีน เอนไซม์ และโมเลกุลดีเอ็นเอ หรืออาร์เอ็นเอ ที่มีความสัมพันธ์กับโรค ทำให้ผลิตภัณฑ์ยาในยุคใหม่นั้น มีความจำเพาะกับโรคมากขึ้น มีความปลอดภัยสูงและมีฤทธิ์ข้างเคียงน้อยลง นอกจากนี้ความผันแปรทางพันธุกรรมในมนุษย์ มีผลตอบสนองต่อยารักษาโรคที่แตกต่างกัน ทำให้เกิดศาสตร์สาขาวิชาใหม่ที่เรียกว่า Pharmacogenomics ซึ่งเป็นการรวมศาสตร์ทางด้าน Pharmacology เข้ากับ genomics เพื่อศึกษารูปแบบของสารพันธุกรรมในแต่ละบุคคล ที่ตอบสนองต่อยารักษาโรค ซึ่งนำไปสู่การรักษาโรคที่มีประสิทธิภาพ และมีความปลอดภัยจากการใช้ยามากขึ้น



ภาพที่ 1-3: สาขาวิชาต่าง ๆ ที่เกี่ยวข้องกับชีวสารสนเทศศาสตร์

1.3.4 ด้านนิติวิทยาศาสตร์ ชีวสารสนเทศศาสตร์ ถูกนำมาประยุกต์ใช้ทางนิติวิทยาศาสตร์ (Forensic science) ด้วยเช่นกัน โดยมีฐานข้อมูลของเครื่องหมายโมเลกุลดีเอ็นเอ Short Tandem Repeat (STR database) สำหรับมนุษย์ การใช้เครื่องมือทางชีวสารสนเทศ เพื่อการวิเคราะห์ความสัมพันธ์ทางพันธุกรรม วิเคราะห์ความน่าเชื่อถือในการตรวจสอบความสัมพันธ์ทางสายเลือด หรือตรวจสอบ ดีเอ็นเอที่เกี่ยวข้องกับคดีอาชญากรรม กฎหมาย และความยุติธรรม เป็นต้น (Bianchi and Lio, 2007)

1.3.5 ด้านเทคโนโลยีชีวภาพ จากความรู้ทางชีวสารสนเทศศาสตร์ ที่มีการค้นหา เปรียบเทียบโครงสร้างจีโนมิกส์ รวมถึงการวิเคราะห์หน้าที่การทำงานของยีนและโปรตีน ถูกนำมาประยุกต์ใช้ในด้านเทคโนโลยีชีวภาพ (Biotechnology) โดยมีความเกี่ยวข้องกับศาสตร์ทางด้านวิทยาศาสตร์พื้นฐาน และวิทยาศาสตร์ประยุกต์ในหลายสาขาวิชา ชีวสารสนเทศในด้านเทคโนโลยีชีวภาพถูกนำมาใช้ในการค้นหายีน และโปรตีนเป้าหมาย เพื่อนำมาสร้างนวัตกรรมใหม่ เพื่อก่อให้เกิดประโยชน์กับมนุษย์ ทั้งทางด้านสุขภาพ (ผลิตยา วัคซีน ชุดตรวจสอบโรค) การผลิตอาหารให้เพียงพอกับความต้องการของประชากรโลก รวมถึงถูกนำมาประยุกต์ใช้กับการรักษาสิ่งแวดล้อม และการผลิตพลังงานชีวมวล อีกด้วยเช่นกัน

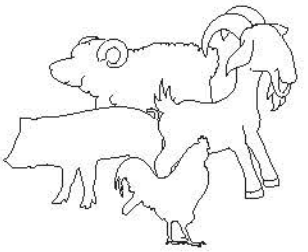
1.3.6 ด้านการเกษตร สำหรับการใช้ชีวสารสนเทศในด้านการเกษตร สามารถแบ่งออกเป็น 2 กลุ่ม คือ ทางด้านพืช และสัตว์ สำหรับทางด้านพืช มีการศึกษาจีโนมของพืช เพื่อใช้สำหรับการปรับปรุงคุณภาพ คุณค่าทางโภชนาการ ความต้านทานต่อโรค และแมลงศัตรูพืช รวมถึงความต้านทานต่อสภาพแวดล้อมที่ไม่เหมาะสม เช่น ทนต่อสภาพความแห้งแล้ง ทนต่อดินเป็นกรด และทนต่อสภาพดินเค็ม เป็นต้น สำหรับในสัตว์ มีการศึกษาจีโนมของสัตว์เศรษฐกิจหลายชนิด เพื่อทำให้เกิดความเข้าใจทางชีววิทยาในระดับโมเลกุลของสัตว์เหล่านี้ได้ดียิ่งขึ้น ซึ่งจะช่วยปรับปรุงพันธุ์สัตว์ ให้มี

ผลผลิตเพิ่มสูงขึ้น ทั้งในเชิงปริมาณและคุณภาพ รวมถึงทำให้สัตว์มีสุขภาพแข็งแรง และช่วยลดการใช้สารเคมีและยาปฏิชีวนะของสัตว์เลี้ยงในฟาร์มลง

1.4 ชีวสารสนเทศทางสัตวศาสตร์

ความรู้ทางด้านชีวสารสนเทศศาสตร์ ถูกนำมาใช้ประโยชน์กับสัตว์เลี้ยงในฟาร์ม ทั้งไก่ สุกร โค แพะ และแกะ รวมถึงสัตว์น้ำหลายชนิด โดยเฉพาะ กุ้ง และปลาชนิดต่าง ๆ โดย ชีวสารสนเทศศาสตร์ ถูกนำมาใช้เป็นเครื่องมือช่วยในการศึกษาชีววิทยาในระดับโมเลกุลของสิ่งมีชีวิตดังกล่าว โดยข้อมูลโมเลกุลดีเอ็นเอ ยีน และโปรตีน ทั้งลำดับนิวคลีโอไทด์ โครงสร้างและหน้าที่การทำงานของยีนและโปรตีน ถูกนำไปเชื่อมโยงกับลักษณะปรากฏ โดยเฉพาะลักษณะทางปริมาณ หรือลักษณะที่มีความสำคัญทางเศรษฐกิจ (economically important traits) ได้แก่ ลักษณะสมรรถภาพการผลิต ลักษณะทางการสืบพันธุ์ ลักษณะการให้ผลผลิต และลักษณะสุขภาพ เป็นต้น ซึ่งนำไปสู่การพัฒนาเป็นฐานข้อมูลทางจีโนมของสัตว์เศรษฐกิจ ได้แก่ ฐานข้อมูลลำดับนิวคลีโอไทด์และลำดับกรดอะมิโน แผนที่ทางพันธุกรรมของสัตว์เศรษฐกิจ (Law and Archibald, 2000; Hu et al., 2001; Hamernik and Adelson, 2003; Fadiel et al., 2005) รวมถึงแผนที่ยีนที่ควบคุมลักษณะทางปริมาณ (quantitative trait loci, QTL map) หรือลักษณะสำคัญทางเศรษฐกิจของสัตว์เลี้ยง (Hu et al., 2007) จากองค์ความรู้ดังกล่าวถูกนำไปบูรณาการสำหรับใช้ในการปรับปรุงพันธุ์สัตว์ ซึ่งมีวัตถุประสงค์เพื่อเพิ่มผลผลิตจากสัตว์ให้มีปริมาณและคุณภาพสูงขึ้น โดยตอบสนองความต้องการของผู้บริโภค

ตัวอย่าง

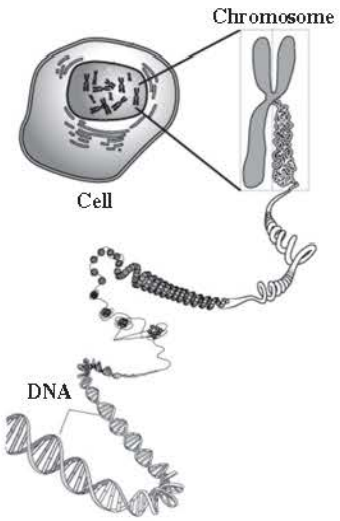


บทที่ 2
รหัสพันธุกรรม

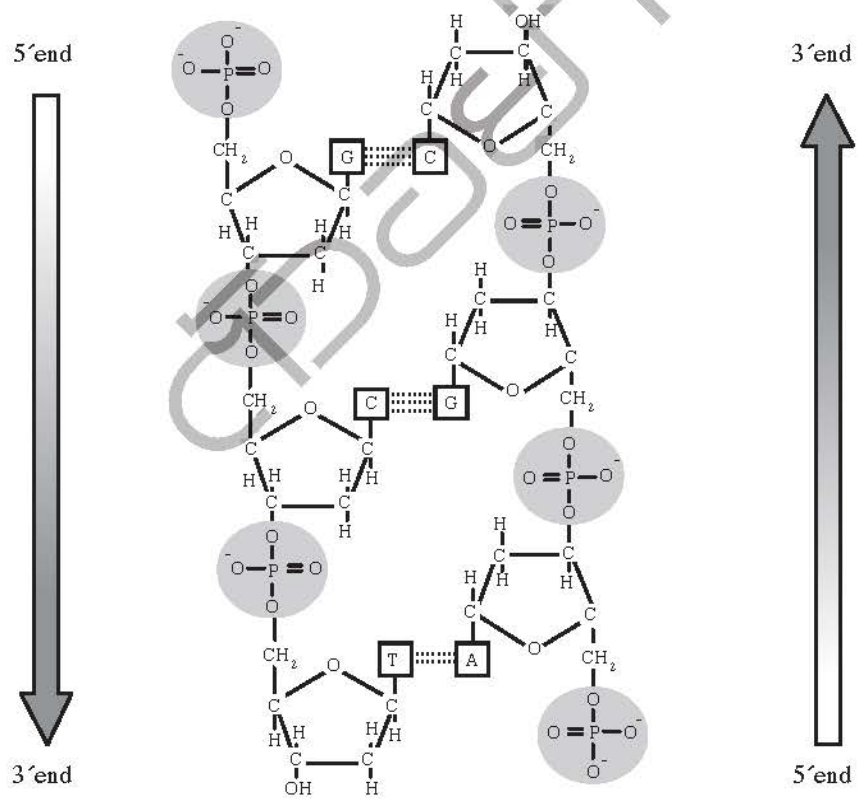
ในบทนี้จะกล่าวถึง ข้อมูลรหัสพันธุกรรม การแปลรหัสพันธุกรรมไปเป็นโปรตีน การถอดรหัสพันธุกรรม และการใช้ประโยชน์จากรหัสพันธุกรรม

2.1 ข้อมูลรหัสพันธุกรรม

รหัสพันธุกรรม (DNA sequence หรือ nucleotide sequence) หรือเรียกสั้น ๆ ว่า ดีเอ็นเอ (deoxyribonucleic acid, DNA) เป็นสารพันธุกรรม ซึ่งเป็นส่วนประกอบที่สำคัญของโครโมโซมที่อยู่ใน nucleus ของเซลล์สิ่งมีชีวิตชั้นสูง (eukaryotic cell) (ภาพที่ 2-1) โดยทำหน้าที่ส่งผ่านข้อมูลทางพันธุกรรมของสิ่งมีชีวิตจากรุ่นหนึ่งไปสู่อีกรุ่นหนึ่ง ดีเอ็นเอจัดเป็นชีวสารที่มีขนาดเล็กในระดับนาโนเมตร ซึ่งมีองค์ประกอบทางเคมีเฉพาะตัว และมีโครงสร้างโมเลกุลเป็นสายคู่ (double helix) ในแต่ละสายของดีเอ็นเอ เกิดจากการเรียงตัวของ นิวคลีโอไทด์ (nucleotide) ที่มีส่วนประกอบย่อย 3 ส่วนคือ (1) โมเลกุลของน้ำตาล pentose (deoxyribose) (2) phosphate และ (3) nitrogenous base โดยที่ nitrogenous base มีโครงสร้างโมเลกุลที่แตกต่างกัน 4 แบบ ประกอบด้วย adenine (A), thymine (T), guanine (G) และ cytosine (C) โครงสร้างโมเลกุล A มีความคล้ายคลึงกับ G ซึ่งจัดอยู่ในกลุ่ม purines ส่วนโครงสร้างโมเลกุล T และ C มีความคล้ายคลึงกัน ซึ่งจัดอยู่ในกลุ่ม pyrimidines การเรียงตัวของ นิวคลีโอไทด์ในแต่ละสายของดีเอ็นเอจะต่อกันด้วยพันธะ phosphodiester bond โดยเกิดจากการเชื่อมต่อระหว่าง phosphate ที่คาร์บอนตำแหน่งที่ 5 ของนิวคลีโอไทด์ตัวหนึ่งกับหมู่ OH ของคาร์บอนตำแหน่งที่ 3 ของนิวคลีโอไทด์อีกตัวหนึ่ง (ภาพที่ 2-2)



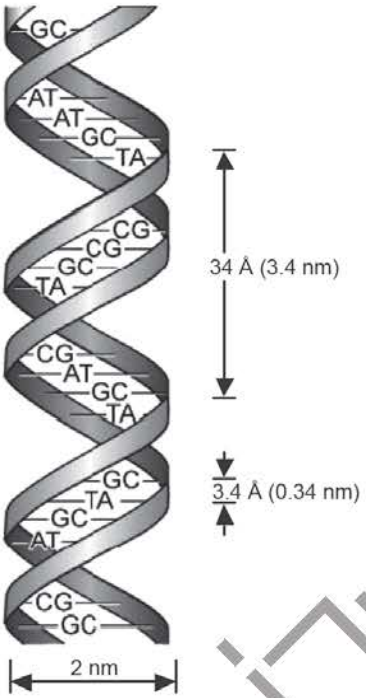
ภาพที่ 2-1: โครงสร้างของเซลล์ โครโมโซม และดีเอ็นเอ (Brooker, 2005)



ภาพที่ 2-2: โครงสร้างโมเลกุลดีเอ็นเอสายคู่ (double-stranded) (ดัดแปลงจาก Brown, 2007)

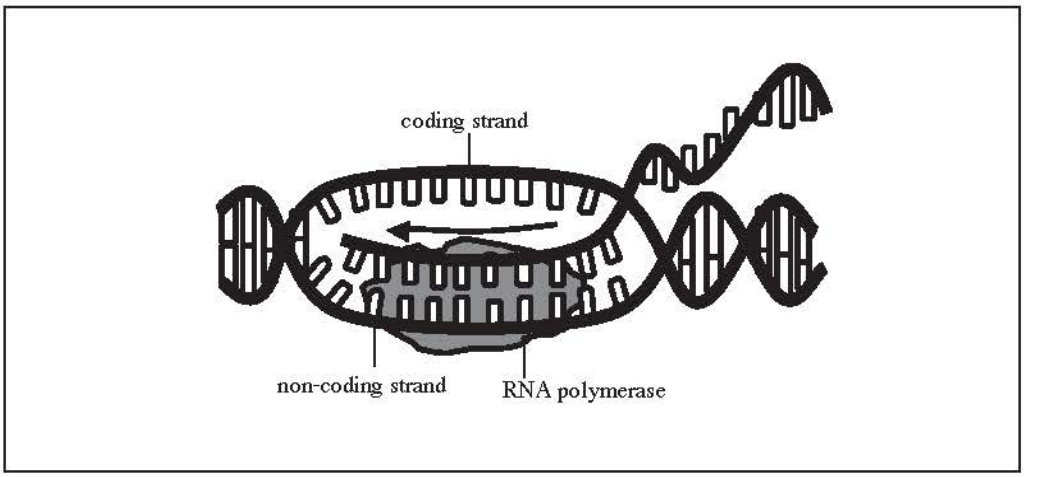
ดังนั้น ลำดับนิวคลีโอไทด์ของโมเลกุลดีเอ็นเอ จึงมีปลายด้านหนึ่งเป็น 5´ phosphate เรียกว่า 5´ end ส่วนด้านปลายตรงกันข้ามจะเป็นหมู่ 3´ OH ซึ่งเรียกว่า 3´ end สำหรับโครงสร้างโมเลกุลของดีเอ็นเอ ที่เป็นสายคู่ (double strand DNA) เกิดจากโครงสร้างโมเลกุลดีเอ็นเอสายเดี่ยวจำนวน 2 สาย เกิดเป็น complementary ซึ่งกันและกัน มาเชื่อมต่อกันด้วยพันธะไฮโดรเจน (hydrogen bond) โดยมี A จับคู่กับ T และ C จับคู่กับ G และมีรูปร่างเป็นเกลียวของ double helix ที่มีเส้นผ่านศูนย์กลางประมาณ 2 นาโนเมตร (nm) ความยาวของดีเอ็นเอ มีหน่วยเรียกว่า base pair (bp) (เช่น 1,000 bp = 1 kb และ 1,000,000 bp = 1 Mb เป็นต้น) ในแต่ละรอบของเกลียวดีเอ็นเอ มีระยะห่างประมาณ 34 อังสตรอม (ภาพที่ 2-3) จากความสัมพันธ์ระหว่างสายรหัสพันธุกรรมที่มีคุณลักษณะ complementary ซึ่งกันและกัน ถ้าเราทราบลำดับนิวคลีโอไทด์เพียงสายใดสายหนึ่ง ก็จะสามารถบอกลำดับนิวคลีโอไทด์ที่เป็นคู่ complementary ได้ (ภาพที่ 2-4) โดยปกติบนสายรหัสพันธุกรรมดีเอ็นเอ จะมีบางบริเวณที่ทำหน้าที่กำหนดการสร้างโปรตีน ซึ่งเรียกบริเวณดังกล่าวว่า ยีน (gene) และมีสายลำดับนิวคลีโอไทด์สายใดสายหนึ่งเพียงด้านเดียวจะทำหน้าที่เป็นรหัสพันธุกรรมกำหนดการสร้างโปรตีน สายดีเอ็นเอดังกล่าว เรียกว่า sense strand หรือ coding strand ส่วนสายดีเอ็นเอด้านตรงกันข้าม ซึ่งเป็น complementary ต่อกัน เรียกว่า anti-sense strand หรือ non-coding strand จะทำหน้าที่เป็นสายดีเอ็นเอแม่แบบสำหรับการสังเคราะห์ mRNA (ภาพที่ 2-5) และ mRNA ที่สังเคราะห์ได้ จะมี sequence คล้ายกับ DNA coding strand ยกเว้นตำแหน่งที่เป็น thymine (T) จะถูกแทนที่ด้วย uracil (U) ซึ่งเรียกว่า primary mRNA โดยลำดับนิวคลีโอไทด์ สามารถแบ่งออกเป็นส่วนเอกซอน (exon หรือ coding sequence) และส่วนอินทรอน (intron หรือ non-coding sequence) หลังจากนั้น primary mRNA จะถูกขจัดเอาส่วนของอินทรอนออก และเหลือเฉพาะส่วนของเอกซอนเท่านั้น ซึ่งเรียกว่า mature mRNA ทำหน้าที่กำหนดการสร้างโปรตีน (open reading frame, ORF) หรือกำหนดคุณลักษณะของสิ่งมีชีวิต (ภาพที่ 2-6)

CPDEMO

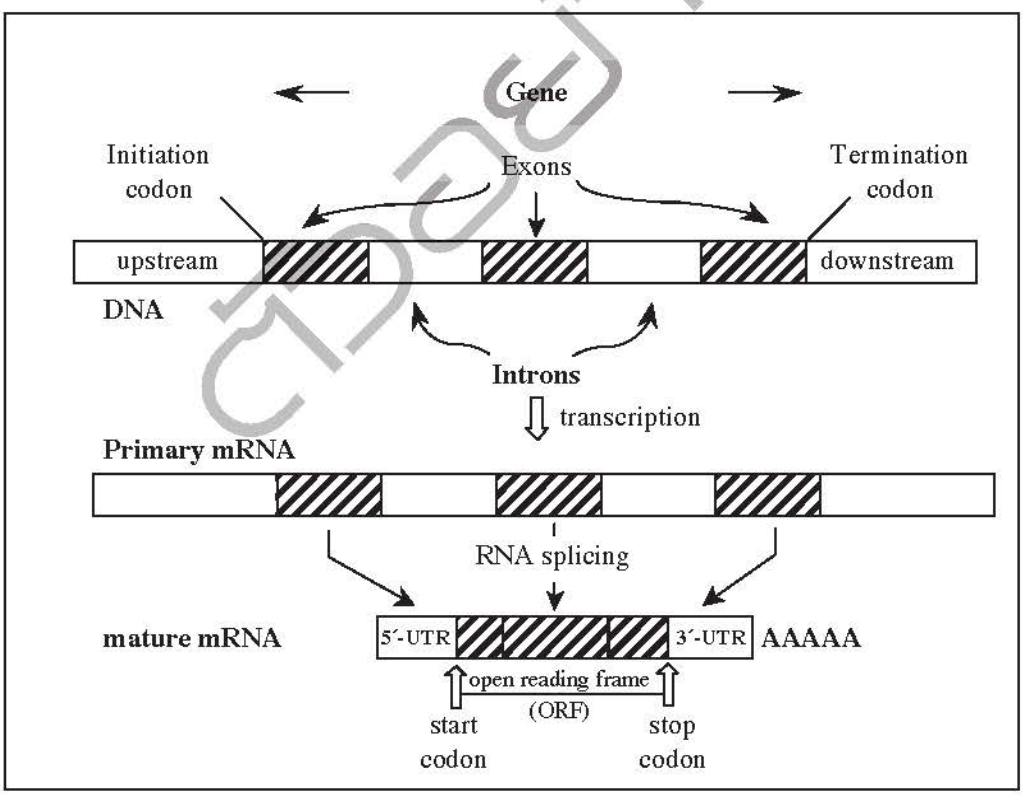


ภาพที่ 2-3: โครงสร้างโมเลกุลดีเอ็นเอ ซึ่งเป็นโมเลกุลสายคู่ (double strand) (White et al., 2007)

ภาพที่ 2-4: คู่สายลำดับนิวคลีโอไทด์ (complementary) ระหว่างสายที่กำหนด (sense strand) และไม่ได้กำหนดการสร้างโปรตีน (anti-sense strand)



ภาพที่ 2-5: การสังเคราะห์ mRNA จากสายดีเอ็นเอแม่แบบ ที่เป็น non-coding strand (ดัดแปลงจาก Snustad and Simmons, 2003)



ภาพที่ 2-6: การสังเคราะห์ mature mRNA (ดัดแปลงจาก Beaumont and Hoare, 2003)

2.2 การแปลรหัสพันธุกรรมไปเป็นโปรตีน

การแปลรหัสพันธุกรรมไปเป็นโปรตีน คือ กระบวนการสังเคราะห์โมเลกุลของโพลีเปปไทด์ (polypeptide) หรือโมเลกุลโปรตีน จากข้อมูลทางพันธุกรรมของโมเลกุล mRNA ซึ่งอาศัยการทำงานของไรโบโซม และ tRNA โดยเรียกกระบวนการดังกล่าวว่า translation ซึ่งเป็นกระบวนการสังเคราะห์โมเลกุลโปรตีนของสิ่งมีชีวิตชั้นสูง เกิดขึ้นในส่วนของไซโตพลาสซึมของเซลล์ โดยรหัสพันธุกรรมที่ประกอบด้วยนิวคลีโอไทด์จำนวน 3 เบส เรียกว่า โคดอน (codon) จะทำหน้าที่เป็นรหัสกำหนดกรดอะมิโน 1 ชนิด โดยที่โคดอน มีทั้งหมด 64 รูปแบบ ซึ่งกำหนดชนิดของกรดอะมิโนที่แตกต่างกัน จำนวน 20 ชนิด (ตารางที่ 2-1 และ 2-2) โดยทั่วไปโมเลกุล mRNA ประกอบด้วยลำดับนิวคลีโอไทด์ 3 ส่วน คือ 5´-untranslated region (5´-UTR), open reading frame (ORF) และ 3´-untranslated region (3´-UTR) (ภาพที่ 2-6) ซึ่งลำดับนิวคลีโอไทด์ในส่วน 5´-UTR และ 3´-UTR ไม่ได้ถูกแปลรหัสไปเป็นโปรตีน สำหรับลำดับนิวคลีโอไทด์ในส่วน ORF คือ ข้อมูลทางพันธุกรรมที่ถูกแปลรหัสไปเป็นโปรตีน ซึ่งเริ่มต้นด้วยโคดอน AUG (start codon) และจบท้ายด้วยโคดอน UAA, UGA หรือ UGA (stop codon)

ตารางที่ 2-1: รหัสพันธุกรรม (genetic code) ที่กำหนดชนิดของกรดอะมิโน (Grifths et al., 2000)

First base	Second base U	Second base C	Second base A	Second base G	Third base
U	UUU - Phe	UCU - Ser	UAU - Tyr	UGU - Cys	U
	UUC - Phe	UCC - Ser	UAC - Tyr	UGC - Cys	C
	UUA - Leu	UCA - Ser	UCA - Stop	UGA - Stop	A
	UUG - Leu	UCG – Ser	UCG - Stop	UGG - Trp	G
C	CUU - Leu	CCU – Pro	CAU - His	CGU - Arg	U
	CUC - Leu	CCC - Pro	CAC - His	CGC - Arg	C
	CUA - Leu	CCA - Pro	CAA - Gln	CGA - Arg	A
	CUG - Leu	CCG - Pro	CAG - Gln	CGG - Arg	G
A	AUU – Ile	ACU – Thr	AAU - Asn	AGU - Ser	U
	AUC - Ile	ACC - Thr	AAC - Asn	AGC - Ser	C
	AUA - Ile	ACA - Thr	AAA - Lys	AGA - Arg	A
	AUG - Met	ACG - Thr	AAG - Lys	AGG - Arg	G
G	GUU - Val	GCU - Ala	GAU - Asp	GGU - Gly	U
	GUC - Val	GCC - Ala	GAC - Asp	GGC - Gly	C
	GUA - Val	GCA - Ala	GAA - Glu	GGA - Gly	A
	GUG - Val	GCG - Ala	GAG - Glu	GGG - Gly	G

ตารางที่ 2-2: รหัสมาตรฐานสำหรับกรดอะมิโน (Claverie and Notredame, 2007)

1-letter code	3-letter code	Amino acid
A	Ala	alanine
C	Cys	cystein
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine
X	Xxx	undetermined amino acid

2.3 การถอดรหัสพันธุกรรม

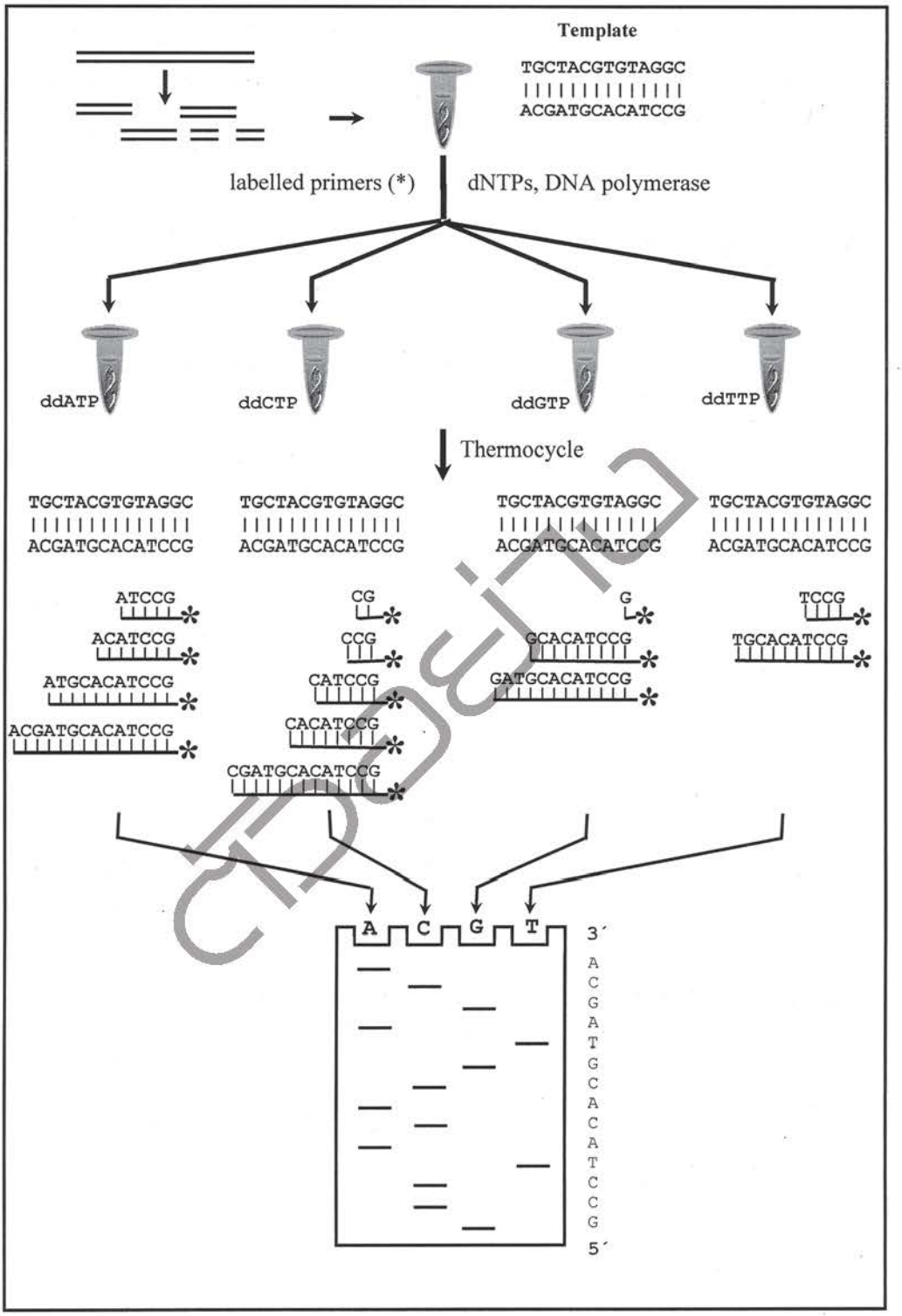
การถอดรหัสทางพันธุกรรม หรือการวิเคราะห์ลำดับนิวคลีโอไทด์ (nucleotide sequence) ของสิ่งมีชีวิตต่าง ๆ ในปัจจุบันนิยมใช้ปฏิกิริยาทางเคมีที่ถูกค้นพบโดย Sanger et al. (1977) ที่เรียกว่า Sanger method หรือ dideoxy or chain-terminating method และใช้เครื่อง automated DNA sequencer (ภาพที่ 2-7) เป็นตัวช่วยอ่านรหัสพันธุกรรม ซึ่งสามารถทำงานได้ในเวลาอันรวดเร็ว หลักการถอดรหัสพันธุกรรม จะอาศัยการเพิ่มปริมาณดีเอ็นเอในหลอดทดลองโดยหลักการ polymerase chain reaction (PCR) แต่มีการยับยั้งการสังเคราะห์สายดีเอ็นเอ ที่ตำแหน่งของเบส A, C, G หรือ T อย่างจำเพาะเจาะจงในปฏิกิริยาดังกล่าว โดยเติมสารเคมี dideoxynucleotides (ddNTPs: ddATP, ddCTP, ddGTP และ ddTTP) เข้าไปยับยั้งปฏิกิริยา นอกจากนี้ยังใช้ไพรเมอร์ (primer)

ภาพที่ 2-7: เครื่อง automated DNA sequencer (http://www.licor.com/)

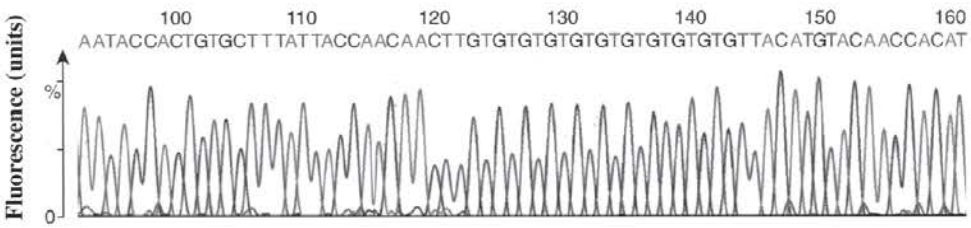
หรือ ddNTPs ที่ถูกติดฉลาก (label) ด้วยสารเรืองแสงฟลูออเรสเซนต์ (fluorescent light) เพื่อความสะดวกต่อการอ่านผลด้วยเครื่อง automated DNA sequencer ซึ่งผลลัพธ์ที่ได้ คือ สายดีเอ็นเอที่มีขนาดสั้นยาวแตกต่างกัน เมื่อนำสายดีเอ็นเอเหล่านี้ ไปแยกขนาดในสนามกระแสไฟฟ้าโดยใช้เครื่อง automated DNA sequencer จะได้รูปแบบของสายดีเอ็นเอ ที่มีการเรียงลำดับตามขนาดความยาวของแถบดีเอ็นเอ ซึ่งมีความสัมพันธ์กับรหัสพันธุกรรมบนสายดีเอ็นเอ (ภาพที่ 2-8) สำหรับตัวอย่างผลการถอดรหัสพันธุกรรมโดยเครื่อง automated DNA sequencer แสดงดังภาพที่ 2-9 อย่างไรก็ตามการถอดรหัสพันธุกรรมอาจได้ผลที่ไม่ชัดเจน จึงมีการกำหนดสัญลักษณ์สากลของลำดับนิวคลีโอไทด์จำนวน 15 รูปแบบ ซึ่งแสดงความหมายดังตารางที่ 2-3

โดยทั่วไปรหัสพันธุกรรมดีเอ็นเอ อาจเป็นส่วนที่กำหนดการสร้างโปรตีน (gene) หรือเป็นส่วนที่ไม่ได้กำหนดการสร้างโปรตีน ทั้งนี้ขึ้นอยู่กับชิ้นส่วนดีเอ็นเอ ที่นำมาถอดรหัสพันธุกรรม ถ้าชิ้นส่วนดีเอ็นเอ มาจาก genomic DNA ก็จะพบส่วนของเอกซอน และอินทรอน ส่วนการถอดรหัสพันธุกรรมของ RNA หรือ cDNA จะพบเฉพาะส่วนของเอกซอนเท่านั้น

ข้อมูลรหัสพันธุกรรมที่ถูกค้นพบโดยนักวิทยาศาสตร์ทั่วโลก จะถูกนำไปเก็บไว้ที่ฐานข้อมูลทางพันธุกรรมของ National Center for Biotechnology Information (NCBI) ที่เรียกว่า GenBank (http://www.ncbi.nlm.nih.gov) ซึ่งเป็นข้อมูลสาธารณะที่ทุกคนสามารถนำไปใช้ประโยชน์ได้โดยไม่ต้องเสียค่าใช้จ่ายแต่อย่างใด



ภาพที่ 2-8: การหาลำดับ DNA sequence โดยวิธีการของ Sanger (ดัดแปลงจาก Grifths et al., 2000)



ภาพที่ 2-9: ผลการถอดรหัสพันธุกรรมแบบใช้สี fluorescence ติดฉลากที่ ddNTP (ดัดแปลงจาก Geldermann, 2005)

ตารางที่ 2-3: สัญลักษณ์ของนิวคลีโอไทด์ (ดัดแปลงจาก Korf et al., 2003)

Symbol	Meaning
A	A (adenine)
C	C (cytosine)
G	G (guanine)
T	T (thymine)
R	A or G (purine)
Y	C or T (pyrimidine)
M	A or C
K	G or T
S	C or G
W	A or T
H	A, C or T not G
B	C, G or T not A
V	A, C or G not T
D	A, G or T not C
N	A, C, G or T

2.4 การใช้ประโยชน์จากข้อมูลรหัสพันธุกรรม

ข้อมูลลำดับนิวคลีโอไทด์ ที่ได้จากการถอดรหัสพันธุกรรม สามารถนำไปใช้ประโยชน์ในทางวิทยาศาสตร์สาขาต่าง ๆ อาทิเช่น ในทางชีววิทยา เพื่อใช้อธิบายปรากฏการณ์ต่าง ๆ ที่เกิดขึ้นในสิ่งมีชีวิต ในด้านการแพทย์มุ่งเน้นการค้นหายีนที่เป็นสาเหตุของโรคที่เกี่ยวข้องกับมนุษย์ เพื่อค้นหาแนวทางการรักษาโรค ให้ตรงกับสาเหตุที่แท้จริง รวมถึงการศึกษาสารพันธุกรรมของเชื้อที่เป็นสาเหตุ

ของโรคในคน การผลิตวัคซีนที่มีประสิทธิภาพสูงในการกระตุ้นภูมิคุ้มกันต่อต้านเชื้อโรค หรือค้นหาแนวทางการผลิตยาที่มีประสิทธิภาพในการรักษาโรค โดยไม่เป็นอันตรายต่อคน หรือไม่เกิดผลข้างเคียง สำหรับงานทางด้านการเกษตร มักจะมุ่งเน้นไปที่การค้นหายีนที่ควบคุมลักษณะการให้ผลผลิตต่าง ๆ หรือยีนที่ต้านทานต่อโรค หรือแมลง รวมถึงการจัดจำแนกสิ่งมีชีวิต เพื่อวัตถุประสงค์ของการอนุรักษ์และการใช้ประโยชน์อย่างยั่งยืนจากแหล่งพันธุกรรม นอกจากนี้ยังรวมถึงการดัดแปลงพันธุกรรมของสิ่งมีชีวิต เพื่อเพิ่มผลผลิต หรือเพิ่มความต้านทานต่อโรค กล่าวโดยสรุปได้ว่าความรู้ทางด้านชีวสารสนเทศศาสตร์ ที่เกี่ยวข้องกับทางด้านการเกษตรนั้น แบ่งออกได้เป็น 3 แนวทาง คือ

(1) การค้นหาจุดกลายพันธุ์ (point mutation) หรือจุดที่มีความผันแปรในโมเลกุลของดีเอ็นเอ (single nucleotide polymorphism) ที่มีผลต่อลักษณะปรากฏของสิ่งมีชีวิต เพื่อสร้างเป็นเครื่องมือตรวจสอบทางพันธุกรรม ใช้สำหรับการปรับปรุงพันธุ์สัตว์ และพันธุ์พืช

(2) การศึกษาในเชิงความหลากหลายทางพันธุกรรม (genetics diversity) ของสัตว์ พืช และจุลินทรีย์ เพื่อจัดจำแนกกลุ่ม เปรียบเทียบความสัมพันธ์ทางพันธุกรรมของสิ่งมีชีวิตในแต่ละสายพันธุ์ หรือจากแหล่งต่าง ๆ

(3) การดัดแปลงพันธุกรรมของสิ่งมีชีวิตให้ได้ลักษณะตามความต้องการ (genetic engineering) รวมถึงการผลิตวัคซีนเพื่อป้องกันโรค

สำนักพิมพ์มหาวิทยาลัยเชียงใหม่ (Chiang Mai University Press)

สำนักพิมพ์มหาวิทยาลัยเชียงใหม่เป็นหน่วยงานหนึ่งภายใต้สังกัดศูนย์บริหารงานวิจัยเริ่มดำเนินการจัดพิมพ์หนังสือตั้งแต่เดือนพฤศจิกายน 2552 เป็นต้นมา การดำเนินงานของสำนักพิมพ์ได้มีการดำเนินงานภายใต้ความเห็นชอบของคณะทำงานจัดตั้งสำนักพิมพ์มหาวิทยาลัยเชียงใหม่ โดยมุ่งเน้นการสนับสนุนและส่งเสริมให้บุคลากรสายวิชาการของมหาวิทยาลัยได้ผลิตงานเขียนในรูปของหนังสือ และสิ่งพิมพ์ เพื่อช่วยสร้างความเข้มแข็งทางวิชาการให้เกิดขึ้น อีกทั้งเป็นการยกระดับมาตรฐานทางงานเขียนของบุคลากรกระตุ้นให้เกิดความตื่นตัวในการสร้างและเผยแพร่ผลงานวิชาการช่วยสร้างภาพลักษณ์และเผยแพร่ชื่อเสียงทางวิชาการของมหาวิทยาลัยอีกทางหนึ่ง

สำนักพิมพ์มหาวิทยาลัยเชียงใหม่จะทำหน้าที่เป็นหน่วยงานกลางในการคัดสรรผลงานเพื่อจัดพิมพ์และเผยแพร่ผลงานเขียนทางวิชาการของบุคลากรสายวิชาการที่เป็นสาระประโยชน์ออกสู่สังคมในปัจจุบันสื่อสิ่งพิมพ์ในการเผยแพร่งานวิชาการมีหลายรูปแบบ และสามารถเผยแพร่ในหลายภาษา ดังนั้นสำนักพิมพ์มหาวิทยาลัยเชียงใหม่มีนโยบายการสนับสนุนการสร้างผลงานคุณภาพออกสู่สังคม โดยกำหนดเป้าหมาย กล่าวคือ

1) การจัดพิมพ์เผยแพร่ผลงานทางวิชาการที่มีคุณภาพของบุคลากรในมหาวิทยาลัย
2) การพัฒนาระบบบริหารจัดการอย่างเป็นระบบ
3) การเผยแพร่ผลงานตีพิมพ์อย่างมีประสิทธิภาพ

ตัวอย่าง

CHIANG MAI
UNIVERSITY PRESS

สำนักพิมพ์มหาวิทยาลัยเชียงใหม่
ศูนย์บริหารงานวิจัย สำนักงานมหาวิทยาลัย มหาวิทยาลัยเชียงใหม่ 50200
โทรศัพท์: 053 94 2676, 053 94 2671-2 โทรสาร: 053 94 3600
E-mail: cmupress.th@gmail.com Website: http://cmupres.cmu.ac.th

ชีวสารสนเทศทางสัตวศาสตร์

ชีวสารสนเทศศาสตร์ เป็นเครื่องมือที่สำคัญในการศึกษาทางด้านพันธุศาสตร์โมเลกุลของสิ่งมีชีวิต หนังสือชีวสารสนเทศทางสัตวศาสตร์เล่มนี้ ได้แนะนำเกี่ยวกับชีวสารสนเทศศาสตร์ที่เกี่ยวข้องในทางสัตวศาสตร์ และได้รวบรวมแหล่งข้อมูลพันธุกรรมที่สำคัญของสัตว์เศรษฐกิจ การจัดข้อมูลทางพันธุกรรม การเปรียบเทียบข้อมูลแบบ Pairwise alignment และ Multiple alignment ฐานข้อมูลจีโนมสัตว์เศรษฐกิจ โดยเฉพาะอย่างยิ่ง ฐานข้อมูล QTL ฐานข้อมูล SNP ฐานข้อมูลจีโนม UCSC และฐานข้อมูลจีโนม Ensembl หนังสือเล่มนี้เหมาะสำหรับนักศึกษาระดับปริญญาโทและปริญญาเอก รวมถึงนักวิชาการและนักวิจัยที่สนใจด้านพันธุศาสตร์โมเลกุล

CHIANG MAI
UNIVERSITY PRESS

ISBN 978-974-672-816-4

ราคา 420 บาท