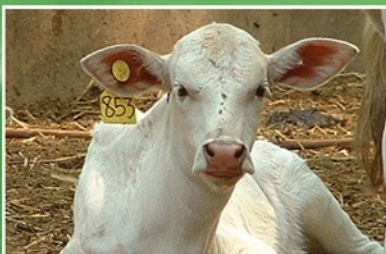


ชีวสารสนเทศศาสตร์



ศุภมิตร เมมฉาย

ภาควิชาสัตวศาสตร์และสัตว์น้ำ

คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่

คำนำ

หนังสือชีวสารสนเทศทางสัตวศาสตร์ (Bioinformatics in Animal Science) เล่มนี้ ผู้เขียน ได้พยายามรวบรวมเอกสาร 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 พงษ์เพียรจันทร์ ครูผู้จุดประกายสร้างความสนใจทางด้านคอมพิวเตอร์ให้แก่ผู้เขียน ผู้เขียนขอกราบขอบพระคุณ Prof. Dr. Siriluck Ponsuksili และ Prof. Dr. Klaus Wimmers ซึ่งเป็นครูผู้แนะนำให้ผู้เขียนรู้จักศาสตร์ของชีวสารสนเทศเป็นครั้งแรก และขอกราบขอบพระคุณ Prof. Dr. Karl Schellander, University of Bonn, Germany ที่ให้การสนับสนุนทางด้านวิชาการเป็นอย่างดีเยี่ยม นอกจากนี้ใคร่ขอขอบคุณกำลังใจและแรงผลักดันจากครอบครัวที่เป็นกำลังใจสำคัญในการเขียนหนังสือเล่มนี้ให้สำเร็จลุล่วงไปได้ด้วยดี ขอขอบคุณภาควิชาสัตวศาสตร์และสัตวน้ำ คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่สนับสนุนและให้โอกาสที่ดีมาโดยตลอด ขอขอบคุณนักศึกษาปริญญาโท และปริญญาเอก ที่เป็นกำลังใจสำคัญในการพิสูจน์อักษรของหนังสือเล่มนี้ให้ถูกต้อง สุดท้ายนี้ขอขอบพระคุณผู้ทรงคุณวุฒิภายนอกทั้ง 2 ท่าน รวมถึง รศ. ดร. จิรยุทธ ไชยจารุณวิช ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ให้ข้อเสนอแนะ และช่วยตรวจสอบแก้ไขต้นฉบับหนังสือเล่มนี้ให้สมบูรณ์อย่างยิ่ง

ผู้เขียนหวังว่า หนังสือเล่มนี้จะเป็นประโยชน์ต่อการเรียนรู้ของนักศึกษา และผู้สนใจ ผู้เขียนยินดีน้อมรับข้อเสนอแนะ และข้อท้วงติงทั้งหมด เพื่อทำให้เนื้อหาวิชาการของหนังสือเล่มนี้ มีความสมบูรณ์ยิ่งขึ้น

ศุภมิตร เมฆฉาย

ตุลาคม 2556

สารบัญ

	หน้า
คำนำ	i
คำนิยาม	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

Bioinformatics in Animal Science

*Bioinformatics in Animal Science

Bioinformatics in Animal Science

Bioinformatics in Animal Science

Bioinformatics in Animal Science

Bioinformatics in Animal Science

Bioinformatics in Animal Science



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

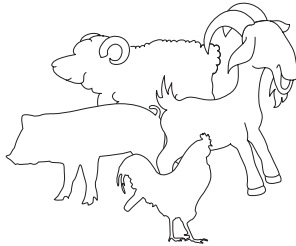


CHIANG MAI
UNIVERSITY PRESS



ราคา 420 บาท

บทที่ 1

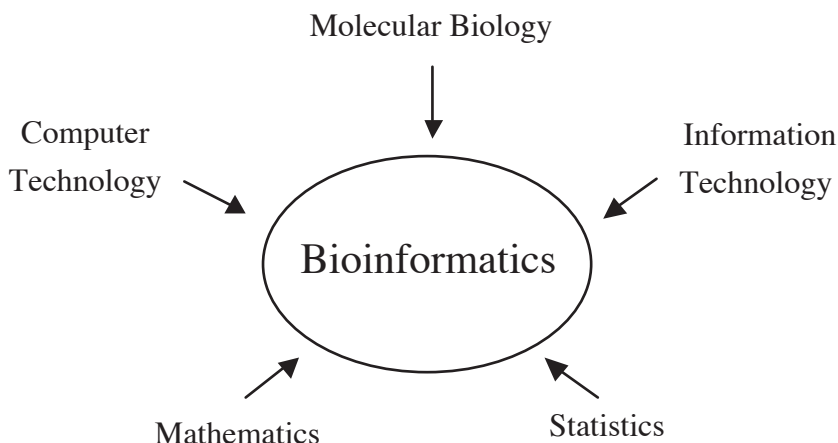


ชีวสารสนเทศศาสตร์

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

1.1 ความหมายของชีวสารสนเทศศาสตร์

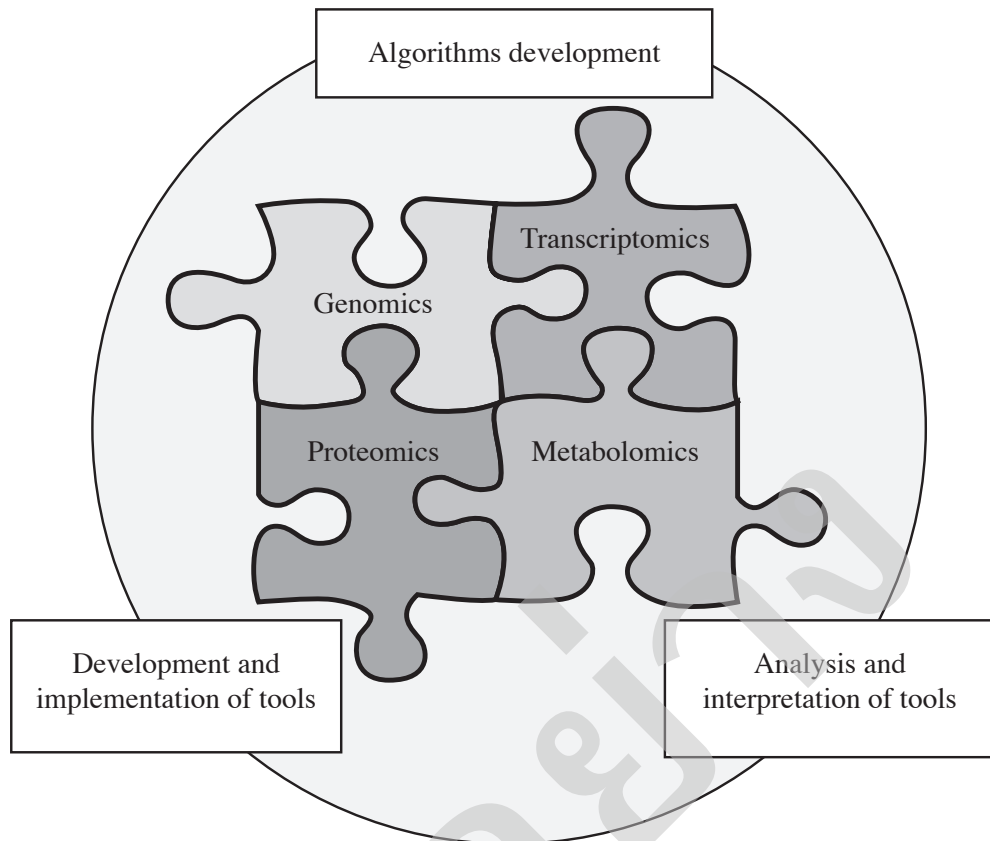
ชีวสารสนเทศ หรือชีวสารสนเทศศาสตร์ (Bioinformatics) เป็นการบูรณาการสหวิทยาการทางด้านคอมพิวเตอร์ เทคโนโลยีทางสารสนเทศ สถิติ และคณิตศาสตร์ เข้ากับองค์ความรู้ทางชีวโมเลกุล (Molecular biology) (ภาพที่ 1-1) เพื่อใช้จัดการและประมวลผลข้อมูลทางชีวโมเลกุลของสิ่งมีชีวิต รวมถึงเก็บรวบรวม วิเคราะห์ สืบค้น เชื่อมโยง และถ่ายโอนข้อมูลทางเครือข่ายคอมพิวเตอร์ (Xiong, 2006; Harisha, 2007; Sharma et al., 2008; Ramsden, 2009) หรือกล่าวอีกนัยหนึ่งได้ว่าชีวสารสนเทศศาสตร์ 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) เมแทบอลอมิกส์ (metabolomics) (ภาพที่ 1-2) โดยที่จีโนมิกส์ เป็นการศึกษาลำดับจีโนม (genome sequence) หรือโครงสร้างของจีโนม (structural genomics) ในขณะที่ ทรานสคริปโตมิกส์ เป็นการศึกษาหน้าที่การทำงานของจีโนม (functional genomics) หรือการแสดงออกของยีน (gene expression) ส่วนโปรตีโอมิกส์ เป็นการศึกษาผลผลิตของยีน ที่มีการแสดงออกมาในเซลล์ หรือในเนื้อเยื่อเป้าหมาย สำหรับเมแทบอลอมิกส์ เป็นการศึกษาสารเมแทบอลิท์ (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 ซึ่งทำให้เกิดความรู้ความเข้าใจ
กลไกในระบบชีววิทยาของสิ่งมีชีวิตมากขึ้น

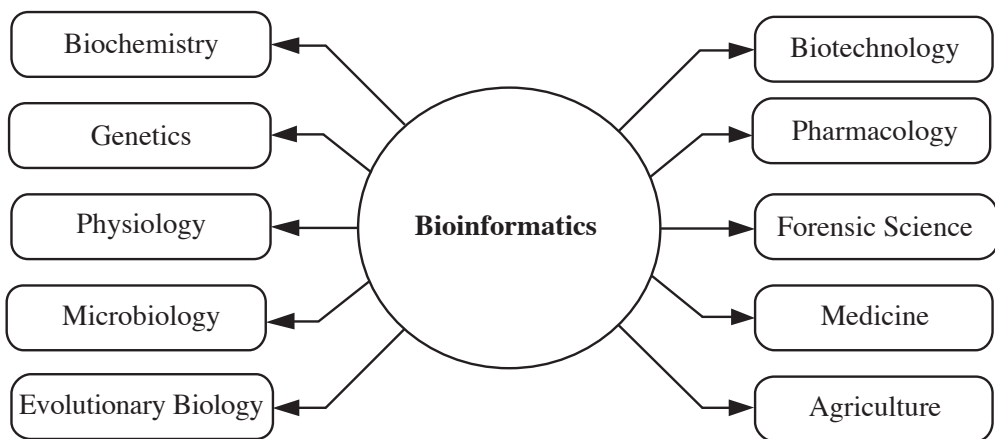
1.3 การประยุกต์ใช้ชีวสารสนเทศศาสตร์

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

1.3.1 ทางด้านวิทยาศาสตร์พื้นฐาน ความรู้ทางด้านชีวสารสนเทศศาสตร์ ถูกนำมาใช้
ใน ทางวิทยาศาสตร์พื้นฐานในสาขาวิชาต่าง ๆ อาทิเช่น ชีวเคมี พันธุศาสตร์ จุลชีววิทยา สรีรวิทยา
และวิวัฒนาการในเชิงชีววิทยา เพื่อสร้างองค์ความรู้พื้นฐานของสิ่งมีชีวิตต่าง ๆ ทั้งพืช สัตว์ และ
จุลินทรีย์ เพื่อนำไปใช้กับวิทยาศาสตร์เชิงประยุกต์ต่อไป

1.3.2 ทางด้านการแพทย์ ผลจากการศึกษาจีโนมของมนุษย์ ทำให้ทราบข้อมูลลำดับ
นิวคลีโอไทด์ของมนุษย์ทั้งหมด โดยชีวสารสนเทศศาสตร์ สามารถนำมาช่วยค้นหายีนที่เป็นสาเหตุ
ของโรคผิดปกติทางพันธุกรรมได้เป็นจำนวนมาก รวมถึงโรคที่มีความซับซ้อน (complex disease)
ซึ่งเกิดจากปฏิสัมพันธ์ระหว่างสารพันธุกรรมและสิ่งแวดล้อม เช่น โรคมะเร็ง โรคเบาหวาน และ
โรคหัวใจ โดยชีวสารสนเทศ สามารถช่วยสืบค้น วิเคราะห์และทำนาย ปัจจัยการเปลี่ยนแปลง
ในระดับเซลล์ ทำให้เข้าใจกลไกทางชีววิทยาโมเลกุลของโรคเหล่านี้มากขึ้น ซึ่งนำไปสู่การค้นพบ
องค์ความรู้ใหม่ สำหรับพัฒนาแนวทางการป้องกัน และรักษาโรคดังกล่าวได้อย่างมีประสิทธิภาพ
มากขึ้น นอกจากนี้ความรู้ทางชีวสารสนเทศศาสตร์ ยังถูกนำมาประยุกต์ใช้ในการผลิตวัคซีนเพื่อป้องกัน
โรคอุบัติใหม่ การรักษาโรคด้วยวิธียีนบำบัด (gene therapy) หรือการตรวจวินิจฉัยโรคในระดับโมเลกุล
(molecular diagnosis) เป็นต้น

1.3.3 ด้านเภสัชศาสตร์ ความรู้ทางชีวสารสนเทศศาสตร์ ถูกนำมาใช้ค้นหาโมเลกุลเป้าหมาย
สำหรับใช้ออกแบบยารักษาโรค ทำให้เกิดการพัฒนายาชนิดใหม่ขึ้นมา โดยอาศัยความรู้พื้นฐานของ
โปรตีน เอนไซม์ และโมเลกุลดีเอ็นเอ หรืออาร์เอ็นเอ ที่มีความสัมพันธ์กับโรค ทำให้ผลิตภัณฑ์ยาใน
ยุคใหม่นั้น มีความจำเพาะกับโรคมากขึ้น มีความปลอดภัยสูงและมีฤทธิ์ข้างเคียงน้อยลง นอกจากนี้
ความผันแปรทางพันธุกรรมในมนุษย์ มีผลตอบสนองต่อยารักษาโรคที่ต่างกัน ทำให้เกิดศาสตร์
สาขาวิชาใหม่ที่เรียกว่า Pharmacogenomics ซึ่งเป็นการรวมศาสตร์ทางด้าน Pharmacology เข้ากับ
genomics เพื่อศึกษารูปแบบของสารพันธุกรรมในแต่ละบุคคล ที่ตอบสนองต่อยารักษาโรค ซึ่งนำไป
สู่การรักษาโรคที่มีประสิทธิภาพ 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 ที่มีการค้นหาเปรียบเทียบโครงสร้างจีโนมิกส์ รวมถึงการวิเคราะห์หน้าที่การทำงานของยีนและโปรตีน ถูกนำมาประยุกต์ใช้ในด้านเทคโนโลยีชีวภาพ (Biotechnology) โดยมีความเกี่ยวข้องกับศาสตร์ทางด้านวิทยาศาสตร์พื้นฐาน และวิทยาศาสตร์ประยุกต์ในหลายสาขาวิชา ชีวสารสนเทศในด้านเทคโนโลยีชีวภาพถูกนำมาใช้ในการค้นหายีน และโปรตีนเป้าหมาย เพื่อนำมาสร้างนวัตกรรมใหม่ เพื่อก่อให้เกิดประโยชน์กับมนุษย์ ทั้งทางด้านสุขภาพ (ผลิตยา วัคซีน ชุดตรวจสอบโรค) การผลิตอาหารให้เพียงพอ กับความต้องการของประชากรโลก รวมถึงถูกนำมาประยุกต์ใช้กับการรักษาสัตว์เลี้ยง และการผลิตพลังงานชีวมวล อีกด้วยเช่นกัน

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

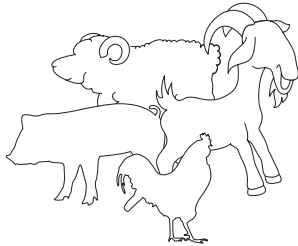
ผลผลิตเพิ่มสูงขึ้น ทั้งในเชิงปริมาณและคุณภาพ รวมถึงทำให้สัตว์มีสุขภาพแข็งแรง และช่วยลดการใช้สารเคมีและยาปฏิชีวนะของสัตว์เลี้ยงในฟาร์มลง

1.4 ชีวสารสนเทศศาสตร์

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 ฐานข้อมูลลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนแผนที่ทางพันธุกรรมของสัตว์เศรษฐกิจ (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) จากองค์ความรู้ดังกล่าวถูกนำไปบูรณาการสำหรับการใช้ในการปรับปรุงพันธุ์สัตว์ ซึ่งมีวัตถุประสงค์เพื่อเพิ่มผลผลิตจากสัตว์ให้มีปริมาณและคุณภาพสูงขึ้น โดยตอบสนองความต้องการของผู้บริโภค

บทที่ 2

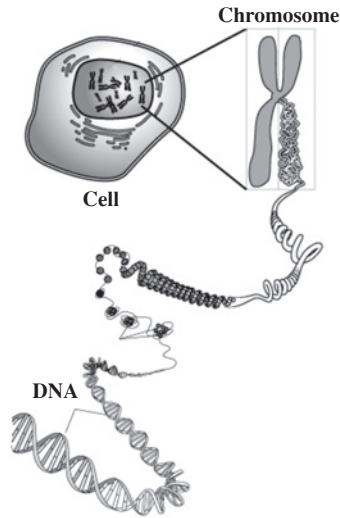
รหัสพันธุกรรม



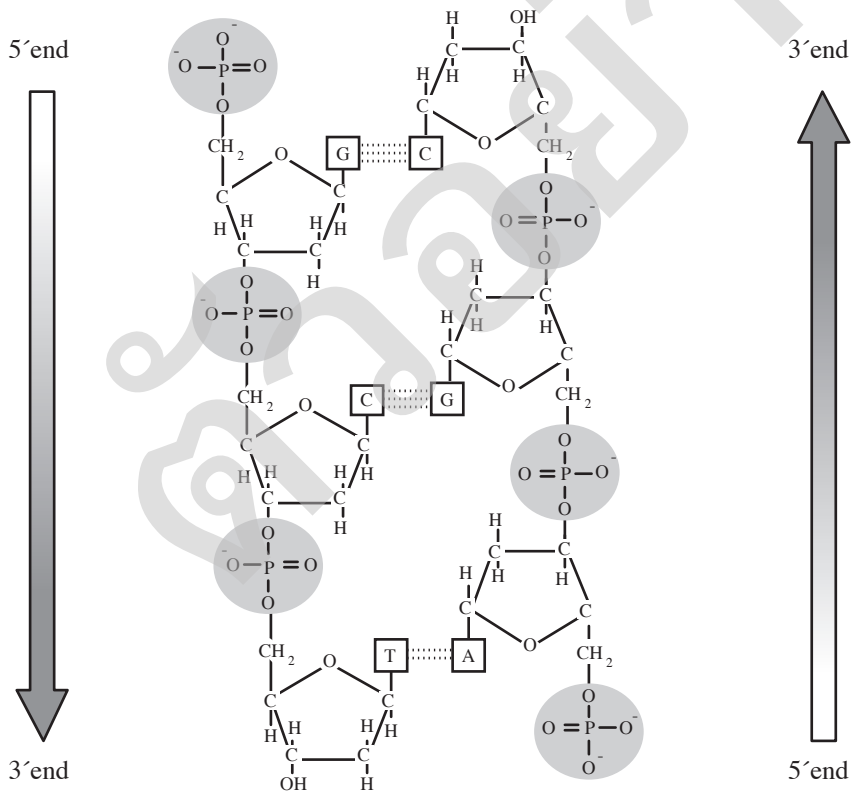
ในบทนี้จะกล่าวถึง ข้อมูลรหัสพันธุกรรม การแปลรหัสพันธุกรรมไปเป็นโปรตีน การถอดรหัสพันธุกรรม และการใช้ประโยชน์จากรหัสพันธุกรรม

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)

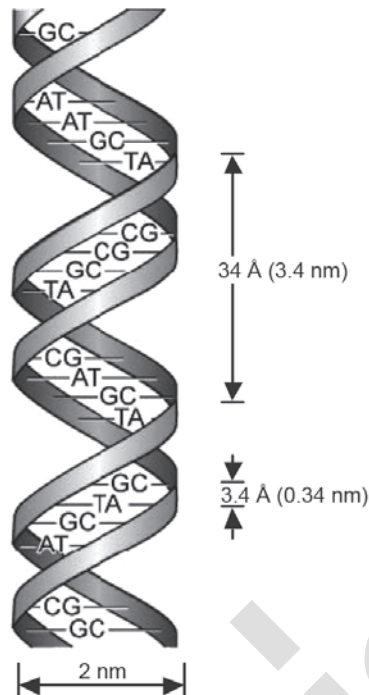


ภาพที่ 2-1: โครงสร้างของเซลล์ โครโมโซม และดีเอ็นเอ (Brooker, 2005)

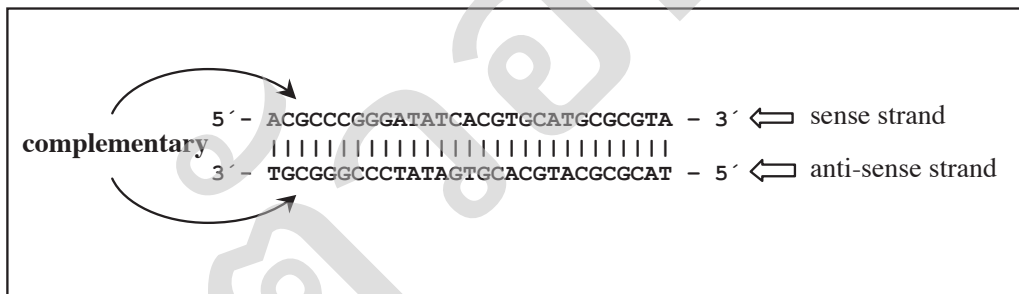


ภาพที่ 2-2: โครงสร้างโมเลกุลดีเอ็นเอสายคู่ (double-stranded) (ดัดแปลงจาก Brown, 2007)

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) และมีสายลำดับนิวคลีโอไทด์สายใดสายหนึ่งเพียงด้านเดียว จะทำหน้าที่เป็นรหัสพันธุกรรมกำหนดการสร้างโปรตีน 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 สามารถแบ่งออกเป็นส่วนของเอกซอน (exon หรือ coding sequence) และส่วนอินตรอน (intron หรือ non-coding sequence) หลังจากนั้น primary mRNA จะถูกจัดเอาส่วนของอินตรอนออก และเหลือเฉพาะส่วนของเอกซอนเท่านั้น ซึ่งเรียกว่า mature mRNA ทำหน้าที่กำหนดการสร้างโปรตีน (open reading frame, ORF) หรือกำหนดคุณลักษณะของสิ่งมีชีวิต (ภาพที่ 2-6)



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



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