

## บทที่ 2

### หลักการและทฤษฎีที่เกี่ยวข้อง

การทำนายยีน มีความรู้พื้นฐาน หลักการ และทฤษฎีที่ต้องทราบ เพื่อให้เกิดความเข้าใจ ในเนื้อหาของงานวิจัยได้ชัดเจนขึ้น ได้แก่ ความหมายทางด้านชีวสารสนเทศ (Bioinformatic) ข้อมูลรหัสพันธุกรรมของดีเอ็นเอ (DNA Data) หน่วยพันธุกรรมยีนออนโทโลยี (Gene Ontology)

#### 2.1 ชีวสารสนเทศ

ชีวสารสนเทศ เป็นการประยุกต์เทคโนโลยีเพื่อการจัดการ และการวิเคราะห์ข้อมูลทางชีววิทยา โดยนำคอมพิวเตอร์มาใช้เพื่อรวบรวม จัดเก็บ วิเคราะห์ และเชื่อมโยงข้อมูลทางชีววิทยาเหล่านั้น หรืออาจกล่าวได้ว่า ชีวสารสนเทศเป็นแนวทางการศึกษาค้นคว้าแบบสหวิทยาการ (Interdisciplinary) ที่เป็นส่วนต่อประสาน (Interface) ระหว่างวิทยาศาสตร์สาขาชีววิทยากับสาขาการคำนวณ เป้าหมายสูงสุดของชีวสารสนเทศคือ เพื่อค้นหาความหมายที่ซ่อนอยู่ในสารสนเทศทางชีววิทยาซึ่งมีปริมาณมาก แล้วนำมาใช้ให้เกิดประโยชน์สูงสุด และเพื่อให้เกิดความเข้าใจอย่างลึกซึ้งต่อชีววิทยาในระดับพื้นฐานของสิ่งมีชีวิต มีการคาดการณ์ว่าความรู้ใหม่ที่ได้จะส่งผลกระทบต่อใหญ่หลวงในหลาย ๆ วงการ เช่น การรักษาโรค เกษตรกรรม สิ่งแวดล้อม พลังงาน และเทคโนโลยีชีวภาพ

สิ่งที่ท้าทายอย่างมากในวงการชีววิทยาระดับโมเลกุล (Molecular biology) ทุกวันนี้ คือการพยายามที่จะทำความเข้าใจคุณค่าของข้อมูลที่ได้มาจากโครงการการจัดลำดับจีโนม (Genome sequencing projects) ซึ่งตามปกติขั้นตอนการวิจัยชีววิทยาระดับโมเลกุลทั้งหมดจะถูกทำให้สำเร็จได้ในห้องปฏิบัติการทดลองเท่านั้น แต่เนื่องจากปริมาณข้อมูลที่เพิ่มขึ้นอย่างมากในยุคจีโนมิกส์ (Genomics) นี้ ทำให้เกิดความจำเป็นที่จะต้องนำคอมพิวเตอร์เข้ามาร่วมในการพัฒนาการวิจัยนี้ด้วยการก่อกำเนิดลำดับเบส (Sequence generation) ซึ่งเป็นหน่วยพื้นฐานของยีน การจัดเก็บลำดับย่อย (Subsequence storage) ของลำดับเบส รวมทั้งการแปล (Interpretation) และการวิเคราะห์ (Analysis) ลำดับเบสเหล่านั้น ทั้งหมดล้วนเป็นงานที่ขึ้นอยู่กับคอมพิวเตอร์โดยตรง อย่างไรก็ตามชีววิทยาในระดับโมเลกุลของสิ่งมีชีวิตหนึ่ง ๆ เป็นเรื่องที่ซับซ้อนมาก จากการวิจัยที่ผ่านมาในระดับที่แตกต่างกัน ทั้งในระดับของกลุ่มยีน (Genome) กลุ่มโปรตีน (Proteome) กลุ่มการถ่ายสำเนา (Transcriptome) และกลุ่มการเผาผลาญพลังงาน (Metabolome) ทำให้ปริมาณข้อมูลจีโนมิกส์

มีการเพิ่มขึ้นอย่างรวดเร็ว สิ่งที่ทำทายเป็นอย่างมากในวงการชีวสารสนเทศทุกวันนี้คือการจัดเก็บข้อมูลที่มีปริมาณมากให้มีประสิทธิภาพสูงสุด ดังนั้นจึงเป็นงานสำคัญที่จะต้องจัดให้มีการเข้าถึงข้อมูลได้โดยง่ายและสามารถเชื่อถือได้ ข้อมูลทั้งหมดที่เก็บรวบรวมมาได้จะต้องถูกนำมาวิเคราะห์เพื่อหาความหมายของข้อมูลก่อนนำไปใช้งานต่อไป แต่ด้วยปริมาณที่มากมายมหาศาลทำให้เป็นการยากที่จะเริ่มแปลความหมายทั้งหมดโดยมนุษย์ ดังนั้นเครื่องมือคอมพิวเตอร์ที่เฉลียวฉลาด จึงต้องถูกพัฒนาขึ้นมา เพื่อให้สามารถสกัดเอาสารสนเทศทางชีววิทยาที่มีความหมายออกมาจากข้อมูลทั้งหมดได้

มีการคาดการณ์เกี่ยวกับเป้าหมายในระยะยาวของการผนวกเข้าด้วยกันของสารสนเทศ เพื่อศึกษาเกี่ยวกับกระบวนการดังกล่าวข้างต้นว่าจะทำให้เราสามารถเรียนรู้ ทำความเข้าใจชีววิทยาของสิ่งมีชีวิตต่าง ๆ ได้อย่างสมบูรณ์

(ออมพิโล มโนรัตน์, 2548)

## 2.2 ข้อมูลรหัสพันธุกรรมของดีเอ็นเอ

สิ่งมีชีวิตทุกสิ่งในโลกล้วนดำรงพันธุ์อยู่ได้ โดยการถ่ายทอดลักษณะทางพันธุกรรม (genetic character) จากรุ่นหนึ่งไปสู่อีกรุ่นหนึ่งต่อไป กันไป ตัวอย่างเช่น มนุษย์เราถ่ายทอดลักษณะทางกายภาพ (สีผม สีผิว สีตา ฯลฯ) จากพ่อแม่ไปสู่ลูก เป็นต้น ลักษณะพันธุกรรมที่สืบทอดต่อกันไปนี้แตกต่างกันไปตามเผ่าพันธุ์ของสิ่งมีชีวิต และถูกควบคุมโดยสิ่งที่เรียกว่ายีน (gene) ที่มีอยู่มากมายในเซลล์ของสิ่งมีชีวิต ยีนเหล่านี้เป็นตัวควบคุมลักษณะทางพันธุกรรมให้แสดงออกมาในสิ่งมีชีวิตแต่ละรุ่น ทำให้การแสดงออกของลักษณะพันธุกรรมที่ได้รับการถ่ายทอดมา มีเพียงบางลักษณะ ซึ่งมีจำนวนน้อยมากเมื่อเทียบกับจำนวนของลักษณะทางพันธุกรรมที่ได้รับการถ่ายทอดมาทั้งหมด ลักษณะทางพันธุกรรมทั้งหมดทั้งที่แสดงออกมาและไม่แสดงออกมา จะถูกถ่ายทอดไปสู่รุ่นต่อไป การถ่ายทอดลักษณะทางพันธุกรรมของสิ่งมีชีวิตนี้จำเป็นต้องอาศัยข่าวสารทางพันธุกรรมของสิ่งมีชีวิตหรือที่เรียกว่า สารพันธุกรรม โดยสำหรับสิ่งมีชีวิตส่วนใหญ่แล้ว จะมีดีเอ็นเอ (DNA) เป็นสารพันธุกรรมหรืออาจกล่าวได้ว่า ดีเอ็นเอเป็นที่เก็บข่าวสารทางพันธุกรรมของสิ่งมีชีวิต ข่าวสารทางพันธุกรรมถูกจัดเก็บในรูปแบบที่เป็นรหัสต่าง ๆ จัดเรียงอยู่บนดีเอ็นเอ รหัสเหล่านี้เรียกว่า รหัสพันธุกรรม (Genetic code) การถ่ายทอดลักษณะทางพันธุกรรมจึงเป็นการถ่ายทอดรหัสพันธุกรรม และเนื่องจากยีนเป็นชุดของรหัสพันธุกรรมที่จัดเรียงอยู่บนดีเอ็นเอ การถ่ายทอดลักษณะทางพันธุกรรมจึงเป็นการถ่ายตดยีนทั้งหมดด้วย

ความสามารถในการควบคุมการแสดงออกของลักษณะทางพันธุกรรมของยีนในสิ่งมีชีวิตแต่ละรุ่น เกี่ยวเนื่องมาจากการที่ยีนมีความเกี่ยวข้องกับการสังเคราะห์โปรตีนของเซลล์ ซึ่งถือว่ามี

ความเกี่ยวข้องกับการแสดงลักษณะและการดำรงชีวิตของสิ่งมีชีวิต ดังนั้นจึงกล่าวได้ว่าดีเอ็นเอเป็นตัวควบคุมลักษณะทางพันธุกรรมของสิ่งมีชีวิตโดยควบคุมการสังเคราะห์โปรตีนนั่นเอง การที่ดีเอ็นเอสามารถควบคุมลักษณะทางพันธุกรรม และการสังเคราะห์โปรตีนของสิ่งมีชีวิตนี้เอง จึงมีผู้พยายามศึกษา และทำความเข้าใจความหมายของรหัสพันธุกรรมที่ดีเอ็นเอเก็บไว้เพื่อเข้าใจลักษณะทางพันธุกรรมของสิ่งมีชีวิต และกลไกการตอบสนองต่อความต้องการ หรือการกระตุ้นจากสิ่งแวดล้อมของเซลล์สิ่งมีชีวิต ในปัจจุบันจึงมีการคิดค้นเครื่องมือเพื่อช่วยในการศึกษา และวิเคราะห์ข่าวสารพันธุกรรมเหล่านี้

(ชโลธร เหลี่ยมวิรัช, 2546)

สารพันธุกรรม หรือดีเอ็นเอ (deoxyribonucleic acid; DNA) เป็นกรดนิวคลีอิก (Nucleic acid) ที่ทำหน้าที่เก็บข้อมูลทางพันธุกรรมของสิ่งมีชีวิต ดีเอ็นเอส่วนใหญ่อยู่ในรูปโครโมโซม (chromosome) วางตัวอยู่ในส่วนนิวเคลียสภายในเซลล์ของสิ่งมีชีวิต ดีเอ็นเอมีหน้าที่สำคัญ 2 ประการ คือ

1. การจำลองตัวเอง (DNA replication) ดีเอ็นเอของสิ่งมีชีวิตมีความสามารถสร้างและจำลองตัวมันเองขณะเกิดกระบวนการแบ่งเซลล์ เพื่อสร้างดีเอ็นเอที่เหมือนเดิมทุกประการให้แก่เซลล์ใหม่

2. การถ่ายทอดข้อมูลผ่านอาร์เอ็นเอ (transcription) ดีเอ็นเอสามารถถอดรหัสเพื่อสร้างเป็นอาร์เอ็นเอ (ribonucleic acid; RNA) อาร์เอ็นเอที่ได้นี้จะทำหน้าที่กำหนด การเรียงตัวของกรดอะมิโนในกระบวนการสังเคราะห์โปรตีน ซึ่งโปรตีนจะถูกนำมาเป็นส่วนประกอบสำคัญในโครงสร้างขององค์ประกอบต่างๆภายในเซลล์ และเป็นสารเร่งปฏิกิริยาทางชีวเคมีหรือเอนไซม์ (enzyme) ในสิ่งมีชีวิต ด้วยหน้าที่ทั้ง 2 ประการของดีเอ็นเอ ทำให้สิ่งมีชีวิตสามารถสืบทอดลักษณะประจำพันธุ์ และดำรงเผ่าพันธุ์อยู่ได้

ดีเอ็นเอประกอบด้วยหน่วยย่อยที่เรียกว่า นิวคลีโอไทด์ (nucleotide) ซึ่งเป็นสารประกอบไนโตรจีนัสเบส (nitrogenous base) แบ่งออกเป็น 2 กลุ่มคือ กลุ่มพิวรีนเบส (purine) ได้แก่ ไทมิน (thymine; T) ไซโทซีน (cytosine; C) และกลุ่มไพริมิดีนเบส (pyrimidine) ได้แก่ อะดีนีน (adenine; A) กัวนีน (guanine; G) โดยสารประกอบไนโตรจีนัสเบสนี้จะรวมตัวกับน้ำตาลดีออกซีไรโบส (deoxyribose sugar) และกรดฟอสฟอริก (phosphoric acid) เป็นนิวคลีโอไทด์อยู่ในดีเอ็นเอ นิวคลีโอไทด์จึงมีอยู่ 4 ชนิดตามชนิดของไนโตรจีนัสเบส คือ อะดีโนซีนไตรฟอสเฟต (adenosine triphosphate; ATP) กัวโนซีนไตรฟอสเฟต (guanosine triphosphate; GTP) ไซโทซีนไตรฟอสเฟต (cytosine triphosphate; CTP) และไทมิดีนไตรฟอสเฟต (thymidine triphosphate; TTP) การ

เรียงลำดับของนิวคลีโอไทด์ ทั้ง 4 ชนิด ส่งผลต่อการเกิดความหลากหลาย และสร้างความแตกต่าง ในลำดับเบสบนสายดีเอ็นเอ ซึ่งมีความจำเพาะในสิ่งมีชีวิตแต่ละชนิด

โครงสร้างของดีเอ็นเอประกอบไปด้วย สายพอลินิวคลีโอไทด์ที่เกิดจากการเชื่อมต่อกัน ของนิวคลีโอไทด์หลายๆหน่วยด้วยพันธะ ฟอสโฟไดเอสเตอร์ โดยเกิดจากสายพอลินิวคลีโอไทด์ จำนวน 2 สายเรียงตัวขนานกันในทิศทางตรงกันข้าม เข้าคู่และพันกันเป็นเกลียวเวียนขวาคลาย บันไดเวียน ที่เรียกว่า ดับเบิลเฮลิคซ์ (doublehelix) การเข้าคู่หรือเข้าจับกันของสายพอลินิวคลีโอ ไทด์ทั้ง 2 สายเกิดจากการเข้าคู่กันระหว่างเบสพิวรีนและเบสไพริมิดีน ด้วยพันธะไฮโดรเจน โดย A ทำการสร้างพันธะจำนวน 2 พันธะเข้าจับกับ T ( $A = T$ ) และ G ทำการสร้างพันธะ จำนวน 3 พันธะ เข้าจับกับ C โดยมีน้ำตาลและหมู่ฟอสเฟตทำหน้าที่เป็นแกนอยู่ด้านนอกของโมเลกุล

(ประสิทธิ์ ผลิตผลการพิมพ์, 2550)

ลำดับเบสที่ต่างกันเมื่อใช้เบสในจำนวนที่แตกต่างกันบนสายพันธุกรรม ทำให้เกิดรหัสที่ แตกต่างกันอย่างมากมาย ตัวอย่างเช่น ถ้าพิจารณาให้เบส 2 โมเลกุลเรียงต่อกันเป็นรหัสแล้วจะได้ จำนวนรหัสที่เกิดจากการจัดเรียงลำดับของเบสเพียง 2 โมเลกุลนี้มากถึง 16 แบบ ( $4^2 = 16$ ) เป็นต้น ในความเป็นจริงแล้ว สายพันธุกรรมประกอบด้วยการจัดเรียงลำดับของเบสหลายโมเลกุลซึ่งอาจมี มากถึงระดับหมื่นโมเลกุล จึงก่อให้เกิดความหลากหลายทางพันธุกรรมของสิ่งมีชีวิตมากมาย

(ชโลธร เหลี่ยมวิรัช, 2546)

### 2.3 หน่วยพันธุกรรม

หน่วยพันธุกรรม หรือยีน (Gene) คือ ดีเอ็นเอส่วนที่เป็นลำดับของรหัสทางพันธุกรรม (Genetic code) สำหรับการสร้างโปรตีนทุกชนิดในเซลล์ของสิ่งมีชีวิต รหัสทางพันธุกรรมเกิดจาก การเรียงลำดับของนิวคลีโอไทด์ทั้ง 4 ชนิด โดยหนึ่งรหัสพันธุกรรมเกิดจากการเรียงตัวของนิวคลี โอไทด์จำนวน 3 นิวคลีโอไทด์ (triplet code) รหัสทางพันธุกรรมหนึ่ง ๆ มีความจำเพาะกับกรดอะมิ โน (amino acid) ที่เป็นองค์ประกอบของโปรตีนเพียง 1 ชนิดจากจำนวนกรดอะมิโนที่มีอยู่ทั้ง 20 ชนิด การเรียงตัวของลำดับ จำนวน และชนิดของกรดอะมิโนที่แตกต่างกันทำให้เกิดเป็นโปรตีน ชนิดต่างๆ

ความแตกต่างของลำดับนิวคลีโอไทด์บนสายดีเอ็นเอของสิ่งมีชีวิตแต่ละชนิด ส่งผลให้เกิด ความแตกต่างกันในระดับรหัสทางพันธุกรรมของสิ่งมีชีวิต ทำให้สิ่งมีชีวิตชนิดเดียวกันแต่ต่างสาย พันธุ์ที่มียีนแตกต่างกันเกิดการสร้างโปรตีนและมีกระบวนการทางชีวเคมีต่างกันออกไป สิ่งมีชีวิต จึงมีความต่างและความหลากหลายในลักษณะทางพันธุกรรมที่ปรากฏออกมา

การค้นหาน้ำที่ของยีนและกลไกการทำงานของยีน สามารถตรวจสอบได้จากผลิตภัณฑ์หรือลักษณะต่างๆที่เป็นการแสดงออกของยีนนั้นในสิ่งมีชีวิตที่ทำการศึกษ โดยเปรียบเทียบระหว่างยีนปกติกับยีนที่ทำงานผิดปกติจากเดิม หรือยีนกลาย (mutated gene) จากการสกัดแยกยีนที่สนใจออกมาจากสิ่งมีชีวิต ทำการชักนำให้เกิดการกลายพันธุ์ และนำมาส่งถ่ายกลับเข้าไปในเซลล์ปกติ แล้วตรวจสอบผลที่เกิดขึ้น ทำให้ทราบว่ายีนนั้นทำงานหรือควบคุมการแสดงออกของลักษณะทางพันธุกรรมอะไร นอกจากนี้ ยังสามารถทำการตัดแปลงยีนให้สร้างผลิตภัณฑ์ตามต้องการได้โดยอาศัยเทคโนโลยีการตัดต่อยีน ด้วยวิธีการตัดแปลงหรือปรับปรุงชิ้นส่วนดีเอ็นเอ หรือยีนให้เป็นไปตามที่ต้องการ แล้วทำการส่งถ่ายยีนเข้าสู่สิ่งมีชีวิตเป้าหมายเพื่อสร้างสิ่งมีชีวิตตัดแปลงพันธุกรรม หรือ GMOs (Genetically Modified Organisms) ต่อไป

ปัจจุบันมีการนำวิธีการด้านเทคโนโลยีชีวภาพต่าง ๆ มาใช้ เพื่อศึกษาทางด้านชีววิทยา ยีน และด้านพันธุกรรม ได้แก่

1. การสกัดแยกดีเอ็นเอออกจากเซลล์
  2. การตัดต่อรวมทั้งการตัดแปลงชิ้นส่วนดีเอ็นเอ
  3. การเพิ่มปริมาณยีนหรือการโคลนยีน (gene cloning)
  4. การเพิ่มขึ้นส่วนดีเอ็นเอที่มีความจำเพาะ จากการทำปฏิกิริยาภายในหลอดทดลองในเครื่องควบคุมอุณหภูมิ
  5. การศึกษาชิ้นส่วนดีเอ็นเอด้วยวิธีแยกขนาดและปริมาณ ผ่านตัวกลางที่เป็นแผ่นวุ้น โดยใช้กระแสไฟฟ้า
  6. การตรวจและพิสูจน์ดีเอ็นเอที่มีการเรียงลำดับเบสที่จำเพาะบนแผ่นเมมเบรน เชื้อพิเศษ (southern blotting)
  7. การหาลำดับเบสบนสายดีเอ็นเอ (DNA sequencing)
  8. การศึกษาความแตกต่างระดับยีน โดยการใช้เครื่องหมายดีเอ็นเอ
  9. การเพาะเลี้ยงเซลล์และการเพาะเลี้ยงเนื้อเยื่อ
  10. การส่งถ่ายยีนเพื่อการเปลี่ยนแปลงลักษณะทางพันธุกรรมของสิ่งมีชีวิตที่ต้องการ
- (สารานุกรมไทยสำหรับเยาวชนฯ เล่มที่ 28, 2547)

## 2.4 จีโนม

จีโนม คือสารพันธุกรรมทั้งหมดของสิ่งมีชีวิตแต่ละชนิด หรือ ยีนทั้งหมดของสิ่งมีชีวิต ซึ่งเป็นสิ่งที่กำหนดและควบคุมลักษณะต่าง ๆ ของสิ่งมีชีวิต ดังนั้นหากทราบว่ายีนทั้งหมดมีวิธีการ



ทำงานอย่างไร ก็สามารปรับปรุ้ง แก้ไข และพัฒนาลักษณะต่าง ๆ ของพืช สัตว์ทุกชนิดให้ดีขึ้นตามที่ต้องการได้

มนุษย์มีดีเอ็นเอทั้งหมด 46 เส้น มีความยาวรวมประมาณ 3,100,000,000 นิวคลีโอไทด์หรือเท่ากับ 3.1 พันล้านคู่เบส (base pair) ซึ่งเป็นที่รวมของสารพันธุกรรมทั้งหมดของมนุษย์ ซึ่งรวมเรียกว่าจีโนม จากโครงการศึกษาจีโนมมนุษย์ (Human Genome Project) ได้มีการหาลำดับเบสบนดีเอ็นเอ (DNA sequencing) และพบว่าบนความยาวทั้งหมดบนเส้นดีเอ็นเอ นั้น มีดีเอ็นเอที่เป็นยีนน้อยกว่า 5% ของสายดีเอ็นเอทั้งหมด มนุษย์มียีนรวมประมาณ 30,000 ถึง 40,000 ยีน แต่ทราบหน้าที่ของยีนทั้งหมดเพียงครึ่งเดียว

นอกจากจีโนมมนุษย์แล้ว ยังมีการศึกษาจีโนมของสิ่งมีชีวิตอื่นๆ อีก เช่น หนู (มีจีโนมขนาดเล็กกว่ามนุษย์เล็กน้อย ขนาดของจีโนมประมาณ 3 พันล้านคู่เบส) แมลงหวี่ (มีจีโนมขนาด 160 ล้านคู่เบส) หนอนตัวกลม (มีจีโนมขนาด 100 ล้านคู่เบส) และแบคทีเรีย E. coli (มีจีโนมขนาด 5 ล้านคู่เบส) โดยมีการหาลำดับคู่เบสได้เกือบหมดแล้ว และนอกจากนี้ยังมีการศึกษาลำดับเบสในพืชอีกด้วย เช่น ใน Arabidopsis และ ข้าว

(ประสิทธิ์ ผลิตผลการพิมพ์, 2550)

## 2.5 ยีนออนโทโลยี

นักวิทยาศาสตร์ส่วนใหญ่เชื่อว่า อาจเป็นไปได้ที่จะมีข้อจำกัดของกลุ่มยีนและโปรตีนที่อยู่ในเซลล์ของสิ่งมีชีวิต การพิจารณาถึงความสอดคล้องกันระหว่างชีววิทยาและข้อมูลเกี่ยวกับยีนและโปรตีนในสิ่งมีชีวิต (Organisms) ต่าง ๆ จะนำไปสู่กระบวนการเปรียบเทียบในระดับจีโนม ความรู้ทางชีววิทยาที่มีบทบาทต่อยีนและโปรตีนในสิ่งมีชีวิตหนึ่ง อาจจะใช้ในการอ้างอิงในอีกสิ่งมีชีวิตหนึ่งได้ ดังนั้นความสามารถในการอ้างอิงดังกล่าวจึงช่วยในการอธิบาย (Annotation) ในสิ่งมีชีวิตหนึ่ง ด้วยคำอธิบายของอีกสิ่งมีชีวิตหนึ่งได้อย่างอัตโนมัติ แต่ปัญหาก็คือฐานข้อมูล หรือองค์ความรู้ที่จะอธิบายสิ่งเหล่านี้มีมากมาย และมีความแตกต่างกันในคำพูดหรือคำศัพท์ที่ใช้ ทั้งที่อยู่ในความหมายเดียวกัน

กลวิธีหนึ่งที่แก้ปัญหานี้คือ การสร้างออนโทโลยี (Ontology) สำหรับเป็นคำอธิบายยีน (Gene Annotation) ร่วมกัน ข้อดีของออนโทโลยีก็คือสามารถที่จะให้คำอธิบายฟังก์ชันการทำงานของยีนในสิ่งมีชีวิตที่ต่างกันโดยอัตโนมัติ ซึ่งออนโทโลยีหนึ่งจะช่วยในการอธิบายคุณสมบัติของยีนเสมอ ตัวอย่างเช่น ยีนตัวหนึ่งอาจให้ผลผลิต (Product) ที่มากกว่าหนึ่ง ผลผลิตของยีนอาจมีคุณสมบัติ ทางชีวเคมี ทางฟิสิกส์ หรือฟังก์ชันการทำงานที่มากกว่าหนึ่งอย่าง และบ่อยครั้งที่ผลผลิต

ของยีนจะอยู่ในส่วนของเซลล์ที่เฉพาะกับมันเอง โครงสร้างหรือคำศัพท์ที่ใช้ในการควบคุมหรือ  
ออนโทโลยีจึงมีภาพที่กว้างกว่าสายพันธุกรรมของนิวคลีโอไทด์ (Sequence of Nucleotides)

ยีนออนโทโลยีเป็นคำศัพท์ที่ใช้ควบคุมคำอธิบายลักษณะของยีน หรือผลผลิตของยีนใน  
สิ่งมีชีวิตใด ๆ มีคุณสมบัติในหลายลักษณะดังนี้

1. เป็นคำศัพท์ควบคุมที่สามารถรับรู้โดยทั่วกัน
2. มีหลายมิติ คือมีมากกว่าหนึ่งออนโทโลยี
3. ใช้อธิบายข้ามสิ่งมีชีวิตได้
4. มีความสัมพันธ์ในหลายลักษณะเช่น ความสัมพันธ์ในลักษณะที่เป็นส่วนประกอบ (part of) หรือ ความสัมพันธ์ในลักษณะที่เป็นอย่างนี้ อย่างนั้น (is a)
5. สามารถแสดงการควบคุมยีนของยีน (regulatory) และความสัมพันธ์ในเชิงปฏิกิริยา (reaction relationship) บางครั้งอาจใช้พาทเวย์ (Pathway) ที่เกี่ยวข้อง
6. แสดงถึงความรู้ที่ยังไม่สมบูรณ์ได้ กรณีเช่น เมื่อไม่ทราบหน้าที่ หรือคุณสมบัติของยีน  
ออนโทโลยีก็สามารถแสดงได้

ยีนออนโทโลยีมีลักษณะของการจัดทำเป็นเครือข่ายของฐานข้อมูลหลายๆเครือข่ายร่วมกัน  
ในการสร้างเป็นยีนออนโทโลยี ฐานข้อมูลเหล่านี้ได้แก่ ยีนแบงก์ (GeneBank) อีเอ็มบีแอล (EMBL)  
ดีดีบีเจ (DDBJ) พีไออาร์ (PIR) มิบส์ (MIPS) วายพีดี (YPD) สวิตซ์-พรีอต (SWIZZ-PROT) เป็นต้น

ยีนออนโทโลยี แบ่งออกเป็น 3 กลุ่มใหญ่ ๆ หรือ 3 ออนโทโลยีซึ่งเป็นอิสระต่อกัน คือ  
กลุ่มกระบวนการทางชีววิทยา (Biological Process) ฟังก์ชันระดับโมเลกุล (Molecular Function)  
และ องค์ประกอบของเซลล์ (Cellular Component) และเมื่อพิจารณาในระดับลึกลงไปอีก ก็จะ  
แสดงให้เห็นถึงออนโทโลยีย่อยที่ใช้อธิบายยีนหรือผลผลิตของยีนได้อีก ซึ่งยีนหรือผลผลิตของยีน  
แต่ละยีนสามารถที่จะอธิบายได้โดยออนโทโลยีหลาย ๆ ออนโทโลยี ดังนั้นจึงอาศัยออนโทโลยีใน  
การอธิบายยีนหรือกลุ่มยีนที่ได้จากการวิเคราะห์ นอกจากนี้ออนโทโลยียังสามารถใช้อธิบาย  
พาทเวย์ในกระบวนการทำงานของยีนในสิ่งมีชีวิตได้อีกด้วย

(Draghici, 2003)