บทที่ 1

บทนำ

ไบโอซิสเตมมาติก (biosystematics) เป็นคำที่มาจากคำ "ไบโอซิสเตมมาตี" (biosystematy) ซึ่ง Camp and Gilly (1943 อ้างถึงใน Solbrig, 1970) ได้เสนอคำ ๆ นี้ โดยมีความหมายดังนี้

An attempt to delimit the natural biotic units [meaning the units of evolution] and to apply these units a system of nomenclature adequate to the task of conveying precise information regarding their defined limits, relationships, variability, and dynamic structure.

โดยความหมายดังกล่าว Solbrig (1970) กล่าวว่าเป็นความพยายามที่จะสร้างระบบการตั้งชื่อซึ่งแสดงให้เห็นความสัมพันธ์โดยเฉพาะความสัมพันธ์ทางสายพันธุ์ระหว่างหน่วยชีวภาพในธรรมชาติ

Lawrence (1967) ให้ความหมายของ "ไบโอซิสเตมมาติก" ว่า "เป็นช่วงหนึ่งของการวิจัยทางพฤกษศาตร์ซึ่งพยายามกำหนดขอบเขตของหน่วยชีวภาพในสภาพธรรมชาติ และทำให้หน่วยเหล่านั้นมีความเด่นชัดซึ่งจำเป็นต้องใช้ข้อมูลจากเคมี นิเวศวิทยา พันธุศาสตร์ เซลล์วิทยา สัณฐานวิทยา การกระจายพันธุ์ และสรีรวิทยา รวมทั้งการสังเกตในธรรมชาติและภายใต้สภาวะควบคุม"

Solbrig (1970) ให้ความหมายของ "ไบโอซิสเตมมาติก" ไว้ว่า "การประยุกต์สาขาวิชาพันธุศาสตร์ เซลล์พันธุศาสตร์ สถิติ และเคมี มาใช้ในการหาคำตอบของการศึกษาปัญหาทางด้านอนุกรมวิธานเพื่อหาคำอธิบายเกี่ยวกับความหลากหลายของสิ่งมีชีวิตในแง่มุมของทฤษฎีวิวัฒนาการ"

Jone and Luchsinger (1987) กล่าวว่า "ไบโอซิสเตมมาติกเป็นการประยุกต์เอาวิธีการทดลอง พันธุศาสตร์ เซลล์วิทยา และการศึกษาประชากรสิ่งมีชีวิตเพื่อนำมาประกอบการศึกษาอนุกรมวิธานโดยเฉพาะการศึกษาในระดับชนิด (species) หรือ ต่ำกว่า (infra-specific level)"

Stace (1989) ให้ความหมายของ "ไบโอซิสเตมมาติก" ไว้ว่า "การศึกษาอนุกรม-วิธานในหัวข้อที่เกี่ยวข้องกับพันธุศาสตร์ เซลล์วิทยา และนิเวศวิทยา รวมถึงการศึกษาในภาค-สนามและในห้องทดลอง"

จากที่กล่าวมาทั้งหมดข้างต้น อาจกล่าวได้ว่า "ไบโอซิสเตมมาติก" คือ "การศึกษาปัญหาทางอนุกรมวิธานเพื่อสร้างระบบการจำแนกสิ่งมีชีวิตซึ่งแสดงความสัมพันธ์ทางสายพันธุ์ของหน่วยทางอนุกรมวิธาน (taxon) ภายในระบบการจำแนกนั้น โดยการใช้ข้อมูลที่ได้จากการศึกษาทางเคมี พันธุศาสตร์ เซลล์พันธุศาสตร์ สัณฐานวิทยา สรีรวิทยา นิเวศวิทยา และการกระจายพันธุ์"

Lawrence (1967) ได้กล่าวถึงวิธีการศึกษาทางไบโอซิสเตมมาติกไว้ว่าประกอบด้วย 3 ขั้นตอนคือ

1. การเก็บตัวอย่างจากประชากรต่าง ๆ และการศึกษาเกี่ยวกับเซลล์วิทยาโดยเฉพาะโครโมโซมในแต่ละประชากรที่เก็บตัวอย่างมา
2. การศึกษาการผสมข้ามระหว่างประชากรต่าง ๆ เพื่อศึกษาความแข็งแรงและความสามารถในการสืบพันธุ์ของลูกผสม
3. การศึกษา homology ของโครโมโซมของลูกผสมเพื่อตรวจความสัมพันธ์ระหว่างประชากรว่ามีมากน้อยเพียงใด

Solbrig (1970) ได้เสนอวิธีการที่นำมาใช้ในการศึกษาไบโอซิสเตมมาติกดังนี้

1. วิธีการทางพันธุศาสตร์ เน้นการศึกษาเกี่ยวกับการผสมข้ามระหว่างประชากรและตรวจลักษณะของลูกผสมว่ามีความสัมพันธ์กับพ่อหรือแม่อย่างไร
2. วิธีการทางเซลล์วิทยา เน้นการศึกษาเกี่ยวกับโครโมโซมในแง่ของจำนวน รูปร่าง และพฤติกรรมของโครโมโซมในขณะที่มีการแบ่งนิวเคลียสแบบไมโอซิส
3. วิธีการทางสถิติ เน้นการเก็บตัวอย่างและการคำนวนหาความสัมพันธ์ระหว่างประชากรโดยพิจารณาค่าสัมประสิทธิ์ของจำนวนลักษณะที่พบในทั้ง 2 ประชากรต่อจำนวนลักษณะทั้งหมดที่นำมาพิจารณา

4. วิธีการทางเคมี เน้นการนำเอาองค์ประกอบทางเคมีของพืช ไม่ว่าจะเป็นเมตาบอไลท์ขั้นที่ 1 เมตาบอไลท์ขั้นที่ 2 หรือโปรตีน เพื่อนำมาใช้ในการอธิบายความสัมพันธ์ระหว่างประชากร

Stace (1989) ได้แนะว่าควรมีการแยกวิธีการศึกษาต่าง ๆ ออกจากการศึกษาทางไบโอซิสเตมมาติกเป็นสาขาย่อย ๆ เพื่อให้สะดวกและง่ายต่อการศึกษา เขาได้ยกตัวอย่างสาขาวิชาไว้พอกล่าวได้ดังนี้

1. numerical taxonomy การศึกษาเกี่ยวกับการประมวลผลจากลักษณะที่นำมาศึกษาโดยอาศัยเทคนิคทางสถิติ

2. chemical taxonomy หรือ chemotaxonomy หรือ chemosystematics เป็นการศึกษาเกี่ยวสารเคมีที่เป็นองค์ประกอบของพืชเพื่อนำข้อมูลที่ได้มาใช้ประกอบกับการศึกษาทางอนุกรมวิธาน

ปัจจุบันได้มีการนำเอาความรู้ทางเคมีในแขนงชีวเคมีเข้ามาช่วยในการศึกษาไบโอซิสเตมมาติก (Hillis and Moritz, 1990) เช่น ใช้ไอโซไซม์ โปรตีน ดีเอ็นเอ (DNA) ทำให้เกิดสาขาใหม่แยกจากสาขาเคมีเดิม เรียกว่า โมเลกูลาซิสเตมมาติก (Molecular systematics)

การศึกษาวิจัยครั้งนี้จะศึกษาไบโอซิสเตมมาติกของโคลงเคลงชน (Melastoma villosum Lodd.) จากประชากรต่าง ๆ ในประเทศไทยโดยเน้นการศึกษาทาง numerical taxonomy และแบบแผนของไอโซไซม์

NUMERICAL TAXONOMY

Numerical taxonomy หมายถึง การจัดกลุ่มโดยวิธีการทางคณิตศาสตร์ของหน่วยทางอนุกรมวิธานบนพื้นฐานของสถานะของลักษณะของหน่วยอนุกรมวิธานนั้น ๆ โดยรวมความถึงการศึกษาความสัมพันธ์ทางสายพันธุ์โดยวิธีการทางสถิติหรือคณิตศาสตร์ ในการศึกษาจะแปลงข้อมูลของลักษณะต่าง ๆ ไปเป็นค่าทางคณิตศาสตร์ ซึ่งทำให้ข้อมูลบางประเภทเช่น ข้อมูลทาง serology หรือ paper chromatography เป็นต้น สามารถนำมาใช้ได้ (Sneath and Sokal, 1973)

ข้อได้เปรียบของ numerical taxonomy นั้น Sneath and Sokal (1973) กล่าวพอสรุปได้ดังนี้

1. numerical taxonomy สามารถนำข้อมูลจากแหล่งต่าง ๆ มาใช้ในการศึกษาได้ อาทิ สัณฐานวิทยา สรีรวิทยา เคมี ลำดับเบสของ DNA ลำดับกรดอะมิโนของโปรตีน เป็นต้น ซึ่งเป็นการยากที่นำข้อมูลดังกล่าวทั้งหมดมาใช้ในการศึกษาอนุกรมวิธานโดยวิธีเดิม

2. ผู้ที่มีความชำนาญด้านอนุกรมวิธานไม่สูงมากนักสามารถทำงานทางด้านนี้ได้โดยใช้วิธีการของ numerical taxonomy

3. การกำหนดสถานะของลักษณะในรูปของตัวเลข สามารถรวบรวมโดยระบบประมวลผลทางอิเลกทรอนิค และสามารถนำไปใช้ในการสร้างคำบรรยาย รูปวิธาน แผนที่และเอกสารอื่น ๆ ที่เกี่ยวข้อง

4. การที่มีข้อมูลเป็นตัวเลข ทำให้มีวิธีการจัดจำแนกหน่วยทางอนุกรมวิธานที่แตกต่างกันได้ดีกว่าและตรวจสอบขอบเขตของหน่วยทางอนุกรมวิธานได้ง่าย ซึ่งจะได้การจัดกลุ่มและรูปวิธานที่ดีกว่า

5. การสร้างตารางข้อมูลที่แน่นอนของ numerical taxonomy บังคับให้ผู้ศึกษาใช้ลักษณะจำนวนมากขึ้นและมีคำบรรยายลักษณะที่ดีกว่า ซึ่งจำเป็นต่อการทำให้คุณภาพของการศึกษาอนุกรมวิธานแบบเดิมดีขึ้น

6. สามารถทำการตรวจสอบซ้ำเกี่ยวกับหลักการของอนุกรมวิธานรวมทั้งปัญหาที่ต้องพิสูจน์ของการจัดกลุ่ม

7. numerical taxonomy ทำให้เกิดการแปลผลใหม่ของความคิดทางชีววิทยาจำนวนมาก

หน่วยทางอนุกรมวิธานที่นำมาใช้ในการศึกษาอาจอยู่ในระดับ (taxonomic rank) ใดก็ได้ ไม่ว่าจะเป็นชนิด (species) สกุล (genus) วงศ์ (family) หรือสูงกว่า แต่เมื่อนำเข้ามาศึกษาด้วย numerical taxonomy จะเรียกหน่วยดังกล่าวว่า operational taxonomic unit (OTU) (Sneath and Sokal, 1973) นอกจากนี้ Sneat and Sokal ยังได้เสนอจำนวนลักษณะที่นำมาศึกษาว่าควรมีจำนวนมากที่สุดเท่าที่จะนำมาใช้ได้ ทั้งนี้ไม่ควรใช้ลักษณะดังต่อไปนี้

1. ลักษณะที่ปราศจากความหมาย ได้แก่ลักษณะที่ไม่แสดงถึงลักษณะประจำของสิ่งมีชีวิตนั้น ๆ ตัวอย่างเช่น ชื่อ จำนวนตัวอย่างแห้ง เป็นต้น ลักษณะที่มีการตอบสนองต่อสิ่งแวดล้อมซึ่งมีการแปรผันสูง และเป็นไปไม่ได้ที่จะตัดสินว่าลักษณะใดเป็นผลจากสภาพแวดล้อมหรือลักษณะใดเป็นผลจากพันธุกรรม ตัวอย่างของลักษณะที่เป็นผลของพันธุกรรมแต่ต้องคัดออก เช่น จำนวนใบใน 1 กิ่ง ซึ่งแม้ว่าจะคงที่ในพืชบางชนิดก็ตาม เป็นต้น อย่างไรก็ตาม ผู้ศึกษาอาจรวมลักษณะที่เป็นผลมาจากสภาพแวดล้อม ถ้าวัตถุประสงค์ของการศึกษาเน้นที่ผลของสภาพแวดล้อมที่มีต่อความสัมพันธ์ของลักษณะที่แสดงออก

2. ลักษณะที่มีความสัมพันธ์กัน จะต้องตัดลักษณะใดลักษณะหนึ่งออก ตัวอย่างเช่น ความยาวของเส้นรอบรูปใบแบบโล่ห์ (peltate shaped) กับรัศมีของใบ ในกรณีนี้ต้องตัดลักษณะใดลักษณะหนึ่งออกไป เป็นต้น

3. ลักษณะที่มีความคงตัวใน OTU ทั้งหมดที่นำมาศึกษา เช่น ถ้าพืชที่ศึกษามี 2 กลุ่ม กลุ่มแรกมีจำนวน appendage 1 อัน แต่อีกกลุ่มมี 2 อัน ในกรณีนี้ควรตัดลักษณะจำนวน appendage ออก เพราะแต่ละกลุ่มมีลักษณะนี้คงที่ เป็นต้น

การเตรียมข้อมูลที่จะใช้ในการวิเคราะห์มักอยู่ในรูปของเมตริกซ์ของลักษณะและ OTU ซึ่งคือ $X_{(ลักษณะ,OTU)}$ ดังนี้

	OTU					
ลักษณะที่	1	2	3	4	...	t
1	X_{11}	X_{12}	X_{13}	X_{14}	...	X_{1t}
2	X_{21}	X_{22}	X_{23}	X_{24}	...	X_{2t}
3	X_{31}	X_{32}	X_{33}	X_{34}	...	X_{3t}
:	:	:	:	:	...	:
n	X_{n1}	X_{n2}	X_{n3}	X_{n4}	...	X_{nt}

อย่างไรก็ตามหากทรานสโพส (transpose) เมตริกซ์ข้างต้นจะได้เมตริกซ์ของ OTU กับลักษณะ $X_{(OTU,ลักษณะ)}$ ซึ่งหากบันทึกข้อมูลในเมตริกซ์แบบนี้จะสามารถทำให้อยู่ในรูปของเมตริกซ์ลักษณะกับ OTU ได้โดยการทรานสโพสกลับ ค่าของ X_{nt} ในเมตริกซ์ข้างต้นสามารถนำมาจากลักษณะต่าง ๆ ของ OTU (Sneath and Sokal, 1973) ได้แก่ สัณฐานวิทยา สรีรวิทยา นิเวศวิทยา การกระจายพันธุ์ พฤติกรรม (ใช้ในการศึกษาด้านสัตววิทยา)

ลักษณะที่นำมาใช้สามารถแบ่งได้เป็น 2 ประเภท (Dunn and Everitt, 1982) คือ

1. Qualitative character คือ ลักษณะที่มีเพียง 2 สถานะ คือ มี หรือ ไม่มี บางครั้งอาจมีมากกว่า 2 สถานะ เช่น สีของกลีบดอก เป็นต้น ลักษณะประเภทนี้ไม่สามารถนำมาศึกษาวิเคราะห์ด้วยวิธีการทางสถิติต่อไปได้ (เช่น ไม่สามารถหาค่าเฉลี่ยได้ เป็นต้น)

2. Quantitative character คือ ลักษณะที่มีความผันแปรในแต่ละ OTU ที่นำมาศึกษา ตัวอย่างเช่น ความยาวของก้านชูอับเรณูและอับเรณู เป็นต้น ลักษณะประเภทนี้สามารถนำมาศึกษาด้วยวิธีการทางสถิติได้ (เช่น หาค่าเฉลี่ย เป็นต้น)

การจำแนกประเภทของสถานะของลักษณะทั้งสองประเภทสามารถจำแนกได้ 6 แบบ (Clifford and Stephenson, 1975) คือ

พลอมลดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. Binary เป็นข้อมูลที่มีเพียง 2 สถานะ คือ มี กับ ไม่มี

2. Disorder multistate เป็นข้อมูลที่มีมากกว่า 2 สถานะ แต่ละสถานะมี rank เท่ากัน เช่น สีของกลีบดอก ซึ่งมีถึง 3 สีคือ แดง เหลือง ขาว เป็นต้น

3. Order multistate เป็นข้อมูลที่มีมากกว่า 2 สถานะ แต่ละสถานะมี rank ไม่เท่ากัน เช่น ความยาวของผล มีได้ตั้งแต่ผลสั้นมาก สั้น ยาว ยาวมาก เป็นต้น

4. Ranked เป็นข้อมูลที่พิจารณาจากหน่วยที่จะศึกษาเพียงหน่วยเดียวแต่อยู่ในสภาพที่แตกต่างกัน มักใช้ในการศึกษาด้านนิเวศวิทยา

5. Meristic เป็นข้อมูลที่เป็นจำนวนเต็มหรือปัดให้เป็นจำนวนเต็ม ตัวอย่าง เช่น จำนวนกลีบดอกใน 1 ดอก จำนวนโอวูลใน 1 ช่องของรังไข่ เป็นต้น

6. Continuous เป็นข้อมูลที่มีค่าต่อเนื่อง ได้แก่ ความยาว ความกว้าง

ในการกำหนดค่า X_{nt} บางครั้งจะพบว่ามีบาง OTU ที่ไม่ได้ศึกษาในบางลักษณะ ซึ่งมักให้ค่า NC (Not Compared) แก่ X_{nt} ค่านั้น และไม่นำมาคำนวนในระหว่างการวิเคราะห์ นอกจากนี้ การให้ค่าแก่ X_{nt} อาจเป็นค่าที่ต่อเนื่อง เช่น ความยาว ความกว้าง ความหนา เป็นต้น หรือค่าของจำนวน เช่น จำนวนดอก จำนวนโอวูล ฯลฯ เป็นต้น ทั้งนี้ไม่จำเป็นต้องแปลงค่าที่ต่อเนื่องหรือค่าจำนวนไปเป็นค่าตามที่กล่าวมาข้างต้นได้ ประโยชน์ของการให้ค่า X_{nt} แบบนี้เหมาะกับการวิเคราะห์ในบางเทคนิค เช่น การวิเคราะห์ปัจจัย (factor analysis) การหาค่าสหสัมพันธ์ของเพียร์สัน (Pearson Product-Moment Correlation Coefficient) เป็นต้น

หลังจากที่ได้ข้อมูลของลักษณะต่าง ๆ ของ OTU ที่ต้องศึกษามาแล้ว การที่หาว่า OTU ที่ i มีความเหมือนหรือใกล้กับ OTU ใดบ้างนั้นสามารถทำได้หลายวิธี ซึ่งจะขอกล่าวเพียงบางวิธีเท่านั้น ซึ่งแยกเป็นประเภทได้ดังนี้ (Sneath and Sokal, 1973; Clifford and Stephenson, 1975; Dunn and Everitt, 1982)

1. Distance Coefficients เป็นการหาระยะห่างหรือความต่างระหว่าง OTU ซึ่งมีหลายวิธี ได้แก่

Average Difference หาค่าความต่างระหว่าง OTU ที่ j และ OTU ที่ k โดยที่

Mean Character Differene (*M.C.D.*), M.C.D. - $\frac{1}{n}\Sigma|X_{ij}-X_{ik}|$

วิธีการหาค่าความต่างวิธีนี้เป็นวิธีการที่ง่ายในการคำนวน ค่า M.C.D. ระหว่าง OTU มีค่าต่ำกว่า Euclidean distance ซึ่งจะกล่าวถึงต่อไป

Taxonomic Distance อาจเรียกอีกชื่อว่า Euclidean distance การคำนวนหาค่าระยะห่างระหว่าง OTU ที่ j และ OTU ที่ k ได้จาก

$$\Delta_{jk} = \sqrt{\Sigma(X_{ij}-X_{ik})^2}$$

และเมื่อลักษณะมาก ๆ มักใช้ค่า d_{jk} แทน Δ_{jk} โดยที่

$$d_{jk} = \sqrt{\frac{\Delta_{jk}^2}{n}}$$

Manhattan หรือ City-Block Metric Distance Coefficient หาค่าความต่างระหว่าง OTU ที่ j และ OTU ที่ k ได้จาก

$$d_{jk} = \Sigma|X_{ij}-X_{jk}|$$

ค่าความต่างโดยวิธีนี้ใช้ในการศึกษาทาง phylogeny กันมาก

2. Association Coefficients เป็นการหาค่าความเหมือนระหว่าง OTU โดยคิดจากสัดส่วนของลักษณะที่ปรากฏในทั้งสอง OTU โดยทั่วไปมักใช้กับข้อมูลที่เป็น two-state โดยที่ทำตาราง 2×2 ของ OTU ที่ j และ ที่ k ดังนี้

		OTU j	
		+ (1)	- (0)
OTU k	+ (1)	a	b
	- (0)	c	d

โดยที่ a คือลักษณะที่พบในทั้งสอง OTU b คือ ลักษณะที่พบใน OTU j แต่ไม่พบใน OTU k
d คือลักษณะที่ไม่พบในทั้งสอง OTU c คือ ลักษณะที่พบใน OTU k แต่ไม่พบใน OTU j

m=a+d u=b+c

วิธีการหาค่าความเหมือนจากข้อมูลเช่นนี้มีหลายวิธีได้แก่

Coefficient of Jaccard (S_j) Sneath (1957) เป็นผู้เสนอ โดยเรียกค่านี้ว่า Similarity มักใช้ในกรณีที่ a มีค่ามาก ๆ วิธีการคำนวนใช้สูตร

$$S_j = \frac{a}{a+u} = \frac{a}{a+b+c}$$

Czekanowski Coefficient (S_c) ใช้ในกรณีที่ a มีค่าน้อย คำนวนจาก

$$S_c = \frac{2a}{2a+b+c}$$

Simple Matching Coefficient (S_{sm}) เป็นวิธีที่ง่ายแต่มีข้อเสียที่นำเอาค่า d มาคำนวนค่าความเหมือนด้วย สูตรการคำนวน คือ

$$S_{sm} = \frac{a+d}{a+b+c+d}$$

3. Correlation Coefficient (r_{jk}) คิดจาก

$$r_{jk} = \frac{\Sigma(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k)}{\sqrt{(X_{ij}-\bar{X}_j)^2(X_{ik}-\bar{X}_k)^2}}$$

ค่า r_{jk} นี้คือ Pearson's Product-Moment Correlation Coefficient

ข้อควรระวังในการใช้ค่า r เพื่อหาความเหมือนระหว่าง OTU คือไม่ควรใช้กับข้อมูลที่พิจารณาถึงขนาด เช่น น้ำหนัก ความยาว เป็นต้น เพราะถ้าขนาดของ OTU ที่ศึกษามีการเพิ่มหรือลดในทิศทางเดียวกัน ดังกรณีตัวอย่าง

	ลักษณะที่				
OTU ที่	1	2	3	4	5
i	10	20	7	10	15
j	5	10	3	5	7
k	30	15	40	30	20

เมื่อหาค่า r_{ij} r_{ik} และ r_{jk} จะพบว่ามีค่าเท่ากันคือ 1 ทั้งที่มีขนาดต่างกันทั้ง 3 OTU อย่างไรก็ตามหากใช้ค่า r เพื่อหาความเหมือนระหว่าง OTU โดยคำนึงถึงรูปร่างแทน สามารถนำมาใช้ได้

การหาค่าความเหมือนระหว่าง OTU จะออกมาในรูปของเมตริกซ์ความเหมือน (S) โดยมีเมตริกซ์ความต่าง (U) ซึ่งได้จาก $U = 1 - S$ เมื่อ $0 \leq s_{ij} \leq 1$ ทั้งเมตริกซ์ความเหมือนและเมตริกซ์ความต่างจะเป็นพื้นฐานสำหรับจัดกลุ่มทางอนุกรมวิธาน วิธีการที่ใช้ในการจัดกลุ่มคือ SAHN (Sequential, Agglomerative, Hierarchic, Nonoverlapping) clustering method มีหลักเกณฑ์ดังนี้

S - มีการคำนวนความเหมือนและความต่างระหว่างกลุ่มใหม่ทุกครั้งภายหลังจากมีการรวมกลุ่มเกิดขึ้น

A - การจัดกลุ่มของ OTU เริ่มจาก OTU ที่แยกกันแล้วค่อยรวม OTU เข้าด้วยกัน

H - การรวมกลุ่มลำดับก่อนมี rank ต่ำกว่าการรวมกลุ่มลำดับหลัง

N - OTU เป็นสมาชิกของกลุ่มใดกลุ่มหนึ่งเท่านั้น

ต่อไปนี้จะขอกล่าวถึงวิธีการที่ numerical taxonomy นำมาใช้ในเพื่อสร้างการจัดกลุ่มดังนี้

Factor analysis

factor analysis เป็นเทคนิคการวิเคราะห์ทางสถิติที่ใช้ในการวิเคราะห์ข้อมูลที่มีตัวแปรหรือลักษณะของพืชตั้งแต่ 2 ลักษณะขึ้นไป โดยการสร้างตัวแปรใหม่ในรูปของสมการเชิงเส้นหลายตัวแปรโดยตัวแปรที่ปรากฏในสมการคือตัวแปรเริ่มต้นทั้งหมด ตัวแปรใหม่ที่ถูกสร้างขึ้นจะเรียก **factor** ซึ่งแต่ละ factor สามารถอธิบายความแปรปรวนของประชากรตัวอย่างได้แตกต่างกัน คือ factor ที่ 1 สามารถอธิบายความแปรปรวนของประชากรตัวอย่างได้มากที่สุด factor ที่ 2 สามารถอธิบายความแปรปรวนของประชากรตัวอย่างได้มากที่สุด ภายหลังจากตัดความแปรปรวนที่อธิบายได้ด้วย factor ที่ 1 ออกไปแล้ว factor ที่ 3 สามารถอธิบายความแปรปรวนของประชากรตัวอย่างได้มากที่สุด ภายหลังจากตัดความแปรปรวนที่อธิบายได้ด้วย factor ที่มีก่อนหน้านี้ (คือ 1 และ 2) ออกไปแล้ว และเป็นไปเช่นนี้ใน factor ต่อ ๆ ไป จำนวน factor ทั้งหมดของประชากรตัวอย่างหนึ่งอาจมีจำนวนไม่เท่ากันขึ้นอยู่กับวิธี

การวิเคราะห์ที่นำมาใช้ อย่างไรก็ตาม ผลรวมของความแปรปรวนที่แต่ละ factor อธิบายได้ไม่ว่าวิธีการวิเคราะห์จะเป็นวิธีใดก็ตามจะเท่ากับความแปรปรวนของประชากรตัวอย่าง

วิธีการวิเคราะห์เพื่อหา factor ของข้อมูลชุดหนึ่งนั้น Linderman, Merenda and Gold (1980) ได้แบ่งเป็น 2 วิธี คือ Principal Component Analysis (PCA) และ Common Factor Analysis (CFA) ข้อแตกต่างของเทคนิคการวิเคราะห์ทั้งสองอยู่ที่จำนวน factor และที่มาของความแปรปรวนของ factor ในวิธี PCA ความแปรปรวนที่แต่ละ factor อธิบายได้หามาจากความแปรปรวนของตัวแปรแต่ละตัวและความแปรปรวนร่วมระหว่างตัวแปร (ซึ่งอาจมีเพียง 2 ตัวแปรหรือมีมากกว่า 2 ตัวแปรก็ได้) ซึ่งโดยวิธีการนี้ factor ที่ได้จะมีจำนวนเท่ากับจำนวนตัวแปร แต่ความแตกต่างระหว่าง factor และตัวแปรเริ่มต้นอยู่ที่ factor แต่ละ factor ไม่มีความสัมพันธ์กัน ในขณะที่ตัวแปรเริ่มต้นมีความสัมพันธ์กัน สำหรับวิธี CFA ความแปรปรวนของแต่ละ factor ได้มาจากความแปรปรวนร่วมระหว่างตัวแปรเริ่มต้น ซึ่งไม่จำเป็นที่ตัวแปรเริ่มต้นทุกตัวจะต้องมีความสัมพันธ์กัน ดังนั้นโดยวิธีการนี้จำนวน factor จึงน้อยกว่าจำนวนตัวแปรเริ่มต้น อย่างไรก็ตาม factor แต่ละ factor ต่างก็ไม่มีความสัมพันธ์กัน จะเห็นได้ว่าโดยหลักการของวิธีการ PCA และ CFA นั้น PCA พยายามทำให้ความแปรปรวนของตัวแปรอย่างน้อย 1 ตัวแปรถูกอธิบายได้มากโดย factor เพียง 1 factor แต่ CFA จะคำนึงถึงความแปรปรวนร่วมระหว่างตัวแปรให้สามารถอธิบายโดย factor ได้มากที่สุด โดยไม่คำนึงถึงความแปรปรวนเนื่องมาจากตัวแปรแต่ละตัว ซึ่งในการศึกษาทางด้าน numerical taxonomy นั้นไม่ค่อยนำมาใช้ แต่มักถูกนำมาใช้ในการการศึกษาด้าน behavioral science (Dums and Everitt, 1982; Clifford and Stephenson, 1975)

การวิเคราะห์ข้อมูลโดยวิธีการ PCA ข้อมูลเริ่มต้นอาจอยู่ในรูปของคะแนนดิบหรือค่าสหสัมพันธ์ระหว่างตัวแปรทุกคู่ก็ได้ ในกรณีที่ข้อมูลเป็นคะแนนดิบจะหาค่าสหสัมพันธ์ระหว่างตัวแปรทุกคู่แล้วนำมาหา principal component ของเมตริกซ์ค่าสหสัมพันธ์ระหว่างตัวแปร ซึ่งเป็นการคำนวนหาค่าไอเกน (eigenvalue) และเวกเตอร์ไอเกน (eigenvector) จากสมการ $(R-\lambda I)v=0$ ค่าไอเกนมีจำนวนเท่ากับจำนวนตัวแปร ค่าไอเกนที่ได้แต่ละค่าแสดงความแปรปรวนของข้อมูลที่สามารถอธิบายได้โดย factor แต่ละ factor และเนื่องจากค่าไอเกนมีจำนวนเท่ากับจำนวนตัวแปร ดังนั้น จำนวน factor จึงเท่ากับจำนวนตัวแปร ค่าไอเกนค่าแรกจะมีค่ามากที่สุด ซึ่งเป็นค่าแสดงความแปรปรวนที่ factor ที่ 1 สามารถอธิบายได้ ค่าไอเกนค่าที่ 2 มีค่ารองลงมาและเป็นค่าแสดงความแปรปรวนที่ factor ที่ 2 สามารถอธิบายได้ภายหลังตัดความแปรปรวนที่ factor ก่อนหน้านี้อธิบายได้ออกไปแล้ว เป็นเช่นนี้จนถึงค่าไอเกนค่าสุดท้าย เมื่อหาค่าไอเกนได้แล้ว ให้นำค่านี้ไปหาเวกเตอร์ไอเกนโดยการแทนค่าไอเกนในสมการ $|R-\lambda I|=0$ ค่าไอเกน 1 ค่า ใช้คำนวนหาเวกเตอร์ไอเกน 1 เวกเตอร์ เวกเตอร์ไอเกนมี

สมาชิกในแต่ละแถวเป็นค่าน้ำหนัก (weight หรือ loading) ของตัวแปรใน factor ที่นำค่าไอเกนมาคิด ดังนั้นจึงมีสมาชิกเท่ากับจำนวนตัวแปรและจำนวนเวกเตอร์ไอเกนเท่ากับจำนวน factor (ซึ่งเท่ากับจำนวนตัวแปร) จะเห็นได้ว่าโดยวิธีการนี้จำนวนตัวแปรมิได้ลดลงเลยแม้ว่าตัวแปรของข้อมูลจะถูกเปลี่ยนไปเป็น factor อย่างไรก็ตาม เราสามารถตัด factor ที่อธิบายความแปรปรวนของข้อมูลได้น้อยออกไปได้โดยเก็บ factor ที่สามารถอธิบายความแปรปรวนของตัวอย่างได้มากไว้ ในการพิจารณาว่าจะตัด factor ใดออกหรือเก็บ factor ใดไว้นั้น ให้ดูจากค่าไอเกน ถ้ามีค่าน้อยกว่า 1 ให้ตัด factor นั้นออกไปและไม่ต้องนำมาพิจารณาอีก (Linderman, Marenda and Gold, 1980) ดังนั้นจึงมีเพียง 4-5 factor แรกหรือน้อยกว่าถูกนำมาพิจารณาสำหรับข้อมูลชุดหนึ่ง ๆ และโดยวิธีการนี้จะเห็นได้ว่าจำนวนตัวแปรใหม่ (factor) ของข้อมูลมีจำนวนน้อยลง การตีความหมายของ factor ให้พิจารณาจากค่าน้ำหนักของตัวแปรตัวเดียวกันในแต่ละ factor (คือค่าของสมาชิกในเวกเตอร์ไอเกน) ซึ่งมักมีค่าสูง ๆ ใน factor หนึ่ง และมีค่าต่ำใน factor ที่เหลือ ตัวแปรใดที่มีค่าน้ำหนักบน factor (factor loading) สูงใน factor หนึ่งแสดงว่าตัวแปรดังกล่าวสามารถอธิบายได้โดย factor นั้นได้มาก และใน factor หนึ่ง ๆ จะมีตัวแปรที่มีค่าน้ำหนักสูงอยู่หลายตัวแปร โดยนัยนี้สามารถบอกได้ว่า factor นั้น ๆ ประกอบด้วยตัวแปรอะไรบ้างและมีความหมายอย่างไร เช่น factor ที่ 1 ประกอบด้วยตัวแปรความกว้างและความยาวของใบ กลีบดอก แสดงว่า factor นี้อธิบายความแปรปรวนเกี่ยวกับขนาด factor ที่ 2 จะประกอบด้วยตัวแปรจำนวนดอกย่อยและโอวูล (ovule) แสดงว่า factor นี้อธิบายความแปรปรวนเกี่ยวกับจำนวน เป็นต้น อย่างไรก็ตามถ้าค่าน้ำหนักของตัวแปรหนึ่ง ๆ มีขนาดใกล้เคียงกันในหลาย ๆ factor ทำให้ยากต่อการตีความหมายของ factor จึงมีความจำเป็นที่จะต้องแปลงค่าน้ำหนักของตัวแปรโดยให้ค่าน้ำหนักของตัวแปรแต่ละตัวแปรมีค่าสูงใน factor เพียง 1 factor และมีค่าใกล้ศูนย์ใน factor อื่น ๆ เพื่อสามารถตีความหมายของ factor ง่ายขึ้นโดยไม่ทำให้ความแปรปรวนที่แต่ละ factor อธิบายได้เปลี่ยนแปลงไป (Linderman, Marenda and Gold, 1980) การแปลงค่าน้ำหนักของตัวแปรข้างต้นจัดเป็นวิธีการวิเคราะห์อีกวิธีการหนึ่งซึ่งเรียกว่า "Multiple factor analysis" (Sneath and Sokal, 1973) ผลลัพธ์ที่ได้จากการวิเคราะห์ปัจจัยอีกส่วนหนึ่งคือค่า communality ค่า communality ของแต่ละตัวแปรคิดจากผลรวมของค่าน้ำหนักของตัวแปรในแต่ละ factor (ที่เลือกไว้) ยกกำลังสอง ค่านี้เป็นค่าที่แสดงความแปรปรวนของตัวแปรแต่ละตัวที่สามารถอธิบายได้ด้วย factor ทั้งหมดที่เลือกไว้ ถ้ามีค่ามากแสดงว่าความแปรปรวนของตัวแปรนั้นสามารถอธิบายได้มากด้วย factor ที่เลือกไว้ ในทางกลับกันถ้ามีค่าน้อย แสดงว่าความแปรปรวนของตัวแปรนั้นสามารถอธิบายโดย factor ได้ต่ำ นอกจากนี้ยังสามารถหาความแปรปรวนของตัวแปรที่สามารถอธิบายได้ในแต่ละ factor โดยคิดจากค่าน้ำหนักยกกำลังสอง ถ้าค่านี้มากแสดงว่า factor นั้นอธิบายความแปรปรวนของตัวแปรนั้นได้มาก

ดังนั้นอาจกล่าวโดยสรุปได้ว่า factor analysis เป็นเทคนิคที่นำมาใช้เพื่อลดจำนวนตัวแปรของข้อมูลให้อยู่ในรูปของ factor และให้ความหมายแก่ factor โดยดูจากตัวแปรที่เป็นองค์ประกอบของแต่ละ factor นอกจากนี้ยังมีผลลัพธ์อีกอันหนึ่ง คือ factor score ซึ่งสามารถนำไปใช้ในการวิเคราะห์ขั้นต่อไปหรือนำมาลงจุดบนแกน factor คู่หนึ่ง ๆ เพื่อดูการกระจายตัวของค่าสังเกตของข้อมูล ประโยชน์ที่ได้จากการลงจุด factor score คือ สามารถตรวจพบค่าสังเกตที่มีการกระจายที่ต่างไปจากค่าอื่น ๆ ได้ง่าย (Norusis, 1985)

Principle Coordinate Analysis

เทคนิคนี้มีความแตกต่างจาก Factor analysis มาก และมีความได้เปรียบในการใช้วิเคราะห์มากกว่า กล่าวคือ ในเมตริกซ์ข้อมูลที่นำมาวิเคราะห์ อาจมีประเภทของข้อมูลได้หลาย ๆ แบบ ทั้งยังอนุญาตให้มีค่า NC สำหรับลักษณะที่ไม่สามารถศึกษาได้ในบาง OTU ในขณะที่ใน factor analysis จะต้องเป็นข้อมูลที่ต่อเนื่องหรือจำนวนเท่านั้น สำหรับวิธีการวิเคราะห์มีความแตกต่างไปบ้าง ซึ่งพอกล่าวเป็นขั้น ๆ ได้ดังนี้

1. หาเมตริกซ์ความเหมือน S ของ OTU ที่นำมาศึกษาจากเมตริกซ์ข้อมูล X แล้วหาเมตริกซ์ความต่าง U โดยที่ U = 1 - S

2. หาเมตริกซ์ E โดยที่สมาชิก e_{jk} คำนวนจาก

$$e_{jk} = \frac{U_{jk}^2}{2}$$

3. หาเมตริกซ์ F โดยที่สมาชิก f_{jk} คิดจาก

$$f_{jk} = e_{jk} - \frac{1}{t}\Sigma e_{j\cdot} - \frac{1}{t}\Sigma e_{k} + \frac{1}{t^2}\Sigma\Sigma e_{jk}$$

$$f_{jk} = e_{jk} - \bar{e}_j - \bar{e}_k + \bar{\bar{e}}$$

4. หา principal component ของเมตริกซ์ F จากสมการ $(F-\lambda I)v=0$

อย่างไรก็ตาม ในกรณีที่ข้อมูลเบื้องต้นเป็นค่าต่อเนื่องหรือจำนวน พบว่าการวิเคราะห์ด้วยวิธี Principal Coordinate Analysis ให้ผลเหมือนกับการวิเคราะห์ด้วย PCA (Sneath and Sokal, 1973)

การวิเคราะห์จัดจำแนก (Discriminant analysis)

การวิเคราะห์จัดจำแนกเป็นเทคนิคทางสถิติซึ่งแตกต่างจาก factor analysis กล่าวคือ factor analysis ไม่จำเป็นต้องระบุกลุ่มของค่าสังเกตก่อน แต่การวิเคราะห์จัดจำแนกจะต้องระบุกลุ่มของค่าสังเกตแต่ละค่าก่อนว่าเป็นสมาชิกของกลุ่มใด หลักการของเทคนิคการวิเคราะห์จัดจำแนกนั้นจะสร้างสมการเชิงเส้นหลายตัวแปรขึ้น โดยตัวแปรที่อยู่ในสมการคือตัวแปรที่ได้มาจากข้อมูลที่จะทำการวิเคราะห์ และสมการที่สร้างขึ้นนี้เรียก **สมการจัดจำแนก** ซึ่งจะต้องเป็นสมการที่ทำให้อัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่มมีค่าสูงสุด (Norusis, 1985) จำนวนสมการจัดจำแนกที่มีได้ในการวิเคราะห์ข้อมูลชุดหนึ่ง ๆ จะมีเท่ากับ จำนวนที่น้อยที่สุดระหว่าง p-1 และ n เมื่อ p คือ จำนวนกลุ่มภายในข้อมูลชุดนั้น ๆ และ n คือ จำนวนตัวแปร

ในการวิเคราะห์จัดจำแนก มีวิธีการพิจารณานำตัวแปรเข้ามาวิเคราะห์เพื่อหาสมการจัดจำแนกพอแบ่งได้ 2 แบบ คือ

แบบที่ 1 เป็นวิธีการที่นำเอาตัวแปรทุกตัวแปรมาวิเคราะห์หาสมการจัดจำแนก วิธีการนี้ทุกตัวแปรจะอยู่ในสมการแม้ว่าบางตัวแปรจะมีอำนาจในการจัดจำแนกต่ำ

แบบที่ 2 เป็นวิธีการพิจารณาตัวแปรที่จะนำเข้าหรือตัดออกจากการวิเคราะห์หาสมการจัดจำแนก โดยวิธีการนี้ตัวแปรที่มีอำนาจในการจัดจำแนกสูง ๆ จะอยู่ในสมการ ในขณะที่ตัวแปรที่มีอำนาจในการจัดจำแนกต่ำจะถูกตัดจากออกสมการ ขั้นตอนของวิธีการเลือกตัวแปรโดยวิธีการนี้ ขั้นตอนแรก คือ พิจารณาว่าตัวแปรใดที่มีค่าเฉลี่ยระหว่างกลุ่มไม่แตกต่างกัน ให้ตัดตัวแปรดังกล่าวออก ในขั้นตอนที่ 2 พิจารณาว่าตัวแปรที่เหลืออยู่ตัวแปรใดมีอำนาจในการจัดจำแนกสูง ให้นำตัวแปรดังกล่าวเข้าสู่การวิเคราะห์หาสมการจัดจำแนก และทำเช่นนี้กับตัวแปรตัวต่อไปที่จะนำเข้าสู่การวิเคราะห์ ซึ่งในระหว่างที่วิเคราะห์ หากตัวแปรใดที่มีอำนาจในการจัดจำแนกต่ำในสมการจะถูกคัดออกจากสมการ หรืออาจใช้วิธีการนำตัวแปรที่เหลือเข้าสู่การวิเคราะห์ทั้งหมดแล้วจึงค่อยคัดตัวแปรที่มีอำนาจในการจัดจำแนกต่ำออกภายหลังก็ได้ อย่างไรก็ตามทั้งสองวิธีจะให้ผลลัพธ์ที่เหมือนกัน ซึ่งคือสมการจัดจำแนกที่มีตัวแปรที่มีอำนาจในการจำแนกสูงอยู่ในสมการ

ในการพิจารณาว่าตัวแปรใดมีความสำคัญต่อสมการจัดจำแนกสูงให้พิจารณาจากค่าสัมบูรณ์ของค่าสัมประสิทธิ์ที่เป็นคะแนนมาตราฐานของตัวแปรในสมการ ค่าสัมประสิทธิ์ของตัวแปรใดที่มีค่าสูงแสดงว่าตัวแปรนั้นมีความสำคัญต่อการจัดจำแนกของสมการนั้น ส่วนการทดสอบนัยสำคัญและอำนาจในการจัดจำแนกสมการจัดจำแนกที่ได้มาว่ามีมากน้อยเพียงใด ให้พิจารณาจากค่าไอเกน (eigenvalue) ค่าสหสัมพันธ์คาโนนิคอล (canonical correlation) ค่าวิล์กแลมดา

(Wilks' lambda) และค่าไคสแควร์ (Chi square) ค่าไอเกนเป็นค่าที่บอกสัดส่วนความแปรปรวนของข้อมูลที่สมการนั้นสามารถอธิบายได้โดยเปอร์เซ็นต์ความแปรปรวนที่สมการนั้นอธิบายได้คิดจากผลหารระหว่างค่าไอเกนของสมการนั้นกับผลรวมของค่าไอเกนจากทุกสมการแล้วคูณด้วย 100 ค่าสหสัมพันธ์คาโนนิคอล กล่าวอย่างง่าย คือค่าที่แสดงความสามารถในการทำนายกลุ่มของสมการ ถ้าค่าสหสัมพันธ์คาโนนิคอลเมื่อยกกำลังสองแล้วมีค่าสูงแสดงว่าสามารถใช้สมการนั้นมาทำนายกลุ่มได้ดี ส่วนค่าวิล์กแลมดาเป็นค่าที่ใช้สำหรับทดสอบค่าเฉลี่ยของกลุ่มที่ได้จากสมการจัดจำแนกนั้น ๆ ว่ามีค่าเท่ากันหรือไม่ ค่าวิล์กแลมดาไม่ได้เป็นค่าทดสอบทางสถิติ แต่จะถูกเปลี่ยนไปเป็นค่าไคสแควร์ เมื่อประกอบกับค่า Degree of Freedom (DF) และค่าความเชื่อมั่นที่ระดับหนึ่งแล้วสามารถบอกนัยสำคัญของสมการนั้นได้ ถ้าค่าไคสแควร์มีนัยสำคัญ แสดงว่า ค่าเฉลี่ยของกลุ่มจากสมการนั้นแตกต่างกัน อย่างไรก็ตาม ถึงแม้ว่าค่าไคสแควร์จะมีนัยสำคัญ แต่ไม่ได้หมายความว่าสมการนั้นสามารถใช้ในการแยกกลุ่มได้ดี ดังนั้นจึงต้องใช้ค่าสถิติที่ได้กล่าวมาประกอบกัน อาจกล่าวโดยสรุปได้ว่า สมการที่มีค่าไอเกนและค่าสหสัมพันธ์คาโนนิคอลสูง แต่มีค่าวิล์กแลมดาต่ำ และไคสแควร์มีนัยสำคัญ แสดงว่าสามารถใช้สมการนั้นในการทำนายกลุ่มได้ดี และถ้าการทดสอบอำนาจการจัดจำแนกซึ่งทำได้โดยหาเปอร์เซ็นต์การทำนายกลุ่มของค่าสังเกตจากสมการ ถ้ามีค่ามาก แสดงว่าสมการนั้นมีอำนาจจัดจำแนกสูง ในบางกรณีที่สมการมีนัยสำคัญ แต่มีค่าไอเกน ค่าสหสัมพันธ์คาโนนิคอลค่อนข้างต่ำ และมีค่าวิล์กแลมดาสูง ให้พิจารณาการทำนายกลุ่มจากสมการ ถ้ามีค่าต่ำ แสดงว่าสมการที่ได้มีอำนาจในการจัดจำแนกต่ำ และไม่ควรนำสมการที่หาได้มาใช้ในการทำนายกลุ่มค่าสังเกตที่ไม่ทราบกลุ่มเพราะอาจทำนายกลุ่มผิดได้ นอกจากนี้ยังสามารถทดสอบอำนาจการจัดจำแนกของสมการได้โดยดูว่าค่าสังเกตของแต่ละกลุ่มมีการกระจายอย่างไรโดยใช้คะแนนจัดจำแนกที่หาได้จากการแทนค่าตัวแปรในสมการด้วยค่าของตัวแปรของค่าสังเกตแต่ละค่า จากนั้นนำมาหาความถี่ของช่วงคะแนนจัดจำแนก แล้วมาสร้างเป็นฮิสโตแกรม ถ้าการกระจายของคะแนนจัดจำแนกของแต่ละกลุ่มไม่มีส่วนใดคาบเกี่ยวกันแสดงว่าสมการดังกล่าวใช้ในการจัดจำแนกและทำนายกลุ่มของค่าสังเกตได้ดี วิธีการนี้เหมาะสำหรับกรณีที่มีกลุ่มเพียง 2 กลุ่ม และมักพบเสมอว่ากลุ่มหนึ่งจะมีการกระจายของคะแนนจัดจำแนกเป็นบวก แต่อีกกลุ่มเป็นลบ ซึ่งเมื่อพิจารณาค่าสัมประสิทธิ์ที่เป็นคะแนนมาตราฐานของตัวแปรในสมการ ถ้าตัวแปรใดมีค่าสูงแสดงว่าลักษณะนั้นมีความแตกต่างกันมาก และเป็นลักษณะที่มีความสำคัญต่อสมการจัดจำแนก สำหรับกรณีที่มีกลุ่มมากกว่า 2 กลุ่ม มักจะหาค่าคะแนนจัดจำแนกจากสมการที่ 1 และ 2 (ในกรณีที่มี 3 กลุ่ม) หรือสมการที่ 1 2 และ 3 (ในกรณีที่มีมากกว่า 3 กลุ่ม) แล้วนำค่าคะแนนมาลงจุดบนแกนสมการทั้งสอง (หรือสามแล้วแต่กรณี) จากนั้นพิจารณาว่าคะแนนจัดจำแนกของแต่ละกลุ่มมีการกระจายอย่างไร แกนสมการที่ 1 แยกกลุ่มใดออกจากกันบ้าง ในทำนองเดียวกับแกนสมการที่ 2 และ 3 ถ้าไม่มีคะแนนจัดจำแนกของกลุ่มใดแยกออกจากกันอย่างเด่นชัดแสดงว่าสมการจัดจำแนกที่หาได้มีอำนาจในการจัดจำแนกต่ำ แต่ถ้าพบว่าค่าคะแนนแต่ละกลุ่มมีการกระจายแยกกัน

อย่างเด่นชัดหรือมีส่วนคาบเกี่ยวกันบ้าง ให้ดูว่าแต่ละกลุ่มแยกกันบนแกนสมการใดและตัวแปรใดที่มีผลต่อการแยกกลุ่ม ซึ่งมักพบเสมอว่าจะมีตัวแปรเพียงไม่กี่ตัวที่มีความสำคัญต่อการจัดจำแนกในแต่ละสมการ

โดยสรุปการวิเคราะห์จัดจำแนกเป็นเทคนิคที่ใช้หาว่าแต่ละกลุ่มมีการแยกกันอย่างไร ซึ่งผลลัพธ์ที่ได้จะถูกนำไปใช้ในการทำนายกลุ่มของค่าสังเกตที่ไม่ทราบกลุ่มได้อย่างถูกต้อง ทั้งนี้ความถูกต้องการทำนายกลุ่มขึ้นกับความเชื่อถือได้ของสมการจัดจำแนกที่ได้มา

การวิเคราะห์จัดกลุ่ม (Cluster analysis)

cluster analysis เป็นเทคนิคที่ใช้ในการจัดค่าสังเกตเข้าเป็นกลุ่ม ๆ โดยพิจารณาค่าของตัวแปรในค่าสังเกตทีละคู่ ถ้าผลรวมของผลต่างกำลังสองของตัวแปรเดียวกันระหว่างค่าสังเกตทั้งสองมีค่าน้อยกว่าค่าสังเกตคู่อื่น ๆ จะรวมค่าสังเกตทั้งสองข้างต้นเป็นกลุ่มเดียวกัน แล้วพิจารณาคู่ค่าสังเกตคู่ ๆ ต่อไป โดยนัยนี้ การวิเคราะห์จัดกลุ่มจะคล้ายกับการวิเคราะห์จัดจำแนกในแง่ของการจัดค่าสังเกตให้อยู่ในกลุ่มใดกลุ่มหนึ่ง แต่การวิเคราะห์ทั้งสองมีแนวคิดที่ต่างกัน (Norusis, 1985) คือ การวิเคราะห์จัดจำแนกจะสร้างสมการจัดจำแนกจากค่าสังเกตที่ทราบกลุ่มแน่นอน จากนั้นนำสมการที่ได้ไปใช้ในการทำนายกลุ่มของค่าสังเกตที่ไม่ทราบกลุ่มต่อไป แต่การวิเคราะห์จัดกลุ่มจะไม่คำนึงถึงกลุ่มของค่าสังเกต คือ ไม่ต้องทราบว่าค่าสังเกตแต่ละค่าเป็นสมาชิกของกลุ่มใดก่อนการวิเคราะห์ แต่จะรวมค่าสังเกตเป็นกลุ่ม ๆ โดยเปรียบเทียบค่าตัวแปรในค่าสังเกต ผลลัพธ์ที่ได้คือการแบ่งค่าสังเกตออกเป็นกลุ่ม ๆ ซึ่งผลลัพธ์ที่ได้ไม่สามารถนำไปใช้ในการทำนายกลุ่มของค่าสังเกตที่มาใหม่ได้ และถ้าต้องการจะทราบว่าค่าสังเกตที่มาใหม่เป็นสมาชิกของกลุ่มใดจะต้องทำการวิเคราะห์ใหม่ทั้งหมด ดังนั้นอาจกล่าวได้ว่าการวิเคราะห์จัดจำแนกเป็นการวิเคราะห์เพื่อใช้ในการทำนายกลุ่มของค่าสังเกต แต่การวิเคราะห์จัดกลุ่มเป็นการวิเคราะห์ที่มุ่งตรวจสอบว่าค่าสังเกตเป็นสมาชิกกลุ่มเดียวกันหรือไม่

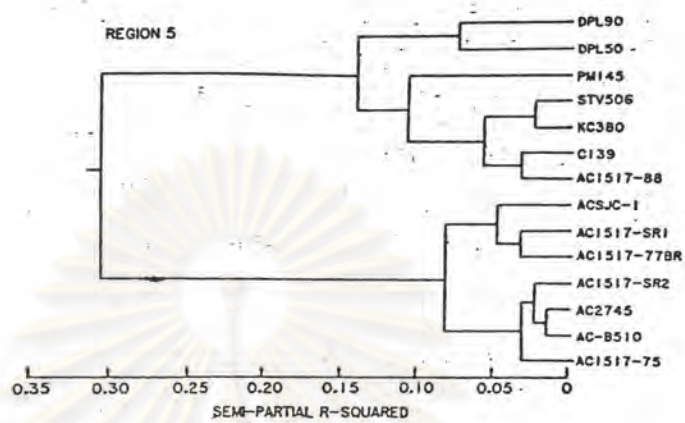
หลักการของการวิเคราะห์ด้วยเทคนิคนี้เริ่มจากการหาระยะห่างระหว่างค่าสังเกตทุกคู่ ซึ่งหาได้จาก $\Sigma(x_{pi}-x_{pj})^2$ เมื่อ p=ตัวแปรที่ 1..p, i และ j = ค่าสังเกตที่ i และ $i \neq j$ จากนั้นเปรียบเทียบว่าระยะห่างของคู่ค่าสังเกตใดมีค่าน้อยที่สุด ให้รวมค่าสังเกตคู่นั้นเป็นกลุ่มเดียวกัน ในขั้นนี้ค่าสังเกตจะเหลือ n-1 ค่า (ค่าสังเกตที่เหลือหลังจากตัดค่าสังเกตออก 1 คู่ รวมกับกลุ่มที่เกิดใหม่จากการรวมกันของค่าสังเกต 2 ค่า ซึ่งเท่ากับ n-2+1=n-1) จากนั้นหาค่าระยะห่างระหว่างค่าสังเกตใหม่แล้วทำการเปรียบเทียบระยะห่างเหมือนข้างต้น ดังนั้นค่าสังเกต

018652

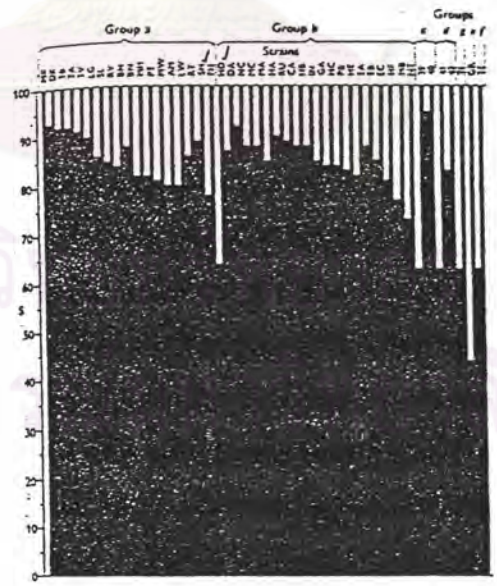
จะลดลงครั้งละ 1 ค่า ผลลัพธ์ที่ได้มักเสนอในรูปเดนโดรแกรม (dendrogram) [ดังแผนภาพที่ 1.1] ซึ่งเป็นแผนภาพแสดงการรวมกลุ่มของค่าสังเกตแต่ละค่าเข้าเป็นกลุ่ม ๆ โดยมีค่าระยะห่างไว้เปรียบเทียบว่าแต่ละค่าสังเกตรวมกันเป็นกลุ่มที่ตำแหน่งใด ค่าระยะห่างที่ให้ไว้จะบอกให้ทราบว่าค่าสังเกตแต่ละค่ามีความเหมือนกันอย่างไรโดยที่ค่าระยะห่างน้อยแสดงว่าเหมือนกันมาก จากเดนโดรแกรมที่ได้ เมื่อกำหนดเส้นที่ตำแหน่งใดตำแหน่งหนึ่งบนค่าระยะห่าง อาจพบว่ามีการแบ่งกลุ่มของค่าสังเกตค่อนข้างเด่นชัด เรียกเส้นที่กำหนดขึ้นมานี้ว่า **เส้นฟีนอน** (phenon line) และเดนโดรแกรมหนึ่ง ๆ อาจมีเส้นฟีนอนได้หลาย ๆ เส้น โดยที่แต่ละเส้นจะอยู่ในตำแหน่งที่มีค่าระยะห่างน้อยกว่าเส้นแรก ๆ อย่างไรก็ตามจำนวนเส้นฟีนอนในเดนโดรแกรมจะมีกี่เส้นนั้นขึ้นอยู่กับว่าในเดนโดรแกรมนั้นควรมีกลุ่มของค่าสังเกตกี่กลุ่มซึ่งผู้ศึกษาจะเป็นผู้ตัดสินเอง นอกจากนี้อาจนำเสนอผลลัพธ์ในรูปอิซิเคิล (icicle) [ดังแผนภาพที่ 1.2] อย่างไรก็ตามมักเสนอผลการวิเคราะห์ในแบบแรกเพราะดูเข้าใจง่ายกว่า

การวิเคราะห์ด้วยเทคนิคนี้ในช่วงแรกที่นำมาใช้ในการศึกษาด้าน numerical taxonomy จะไม่ใช้วิธีการคำนวนระยะห่าง ดังการศึกษาของ Sneath (1957) ในแบคทีเรีย 45 สายพันธุ์ โดยการใช้ตัวแปรซึ่งการทดสอบทางชีวเคมีของแต่ละสายพันธุ์จำนวน 105 การทดสอบแล้วนำมาหาค่าความเหมือน (similarity, S) ของทุกคู่สายพันธุ์จาก $S = n_d/(n_s + n_d)$ โดยที่ n_d คือจำนวนปฏิริยาที่ให้ผล positive ทั้งสองสายพันธุ์ และ n_s คือจำนวนปฏิริยาที่ให้ผล positive ในสายพันธุ์ใดสายพันธุ์หนึ่งในคู่สายพันธุ์นั้น ๆ แล้วทำการเปรียบเทียบว่าคู่ใดมีค่าความเหมือนกันมากให้รวมเป็นกลุ่มเดียวกัน และทำเช่นนี้ไปจนครบ ในการนำเสนอ Sneath ใช้อิซิเคิล (icicle) เพื่อแสดงการรวมกลุ่มของแต่ละสายพันธุ์ ผลการจัดกลุ่มของแบคทีเรียที่นำมาศึกษาพบว่าแบ่งได้ 7 กลุ่ม ซึ่งตรงกับการแบ่งกลุ่มที่เคยศึกษามาก่อนแล้ว

นอกจากเทคนิคการวิเคราะห์ดังกล่าวมาแล้วยังมีเทคนิคอื่น ๆ ที่ได้นำมาใช้ในการศึกษาทางด้าน numerical taxonomy ได้แก่ Canonical Variate Analysis จัดเป็นเทคนิคการวิเคราะห์จัดจำแนก ซึ่ง Clifford and Stephenson (1975) จัดเป็น multiple discriminant analysis วัตถุประสงค์ของเทคนิคนี้นอกจากจะหาสมการจัดจำแนกแล้ว ยังหาความสัมพันธ์ระหว่างกลุ่มของค่าสังเกตที่นำมาศึกษาด้วย ผลลัพธ์ที่ได้จะเรียก canonical variate (คือสมการจัดจำแนก) ความสัมพันธ์ระหว่างกลุ่มดูจากการกระจายของค่าสังเกตบนแกน canonical variate ที่ 1 กับ 2 (และ 3 ขึ้นกับจำนวน canonical variate ที่ได้) ประกอบกับค่าระยะห่างมาฮาลาโนบิส (Mahalanobis, D^2) กลุ่มที่อยู่ห่างกลุ่มอ้างอิงมากจะมีค่า D^2 มากกว่ากลุ่มที่อยู่ใกล้กับกลุ่มอ้างอิง และถ้าทุก ๆ กลุ่มอยู่ใกล้กันมากแสดงว่ากลุ่มที่นำมาศึกษามีความใกล้ชิดกันมาก



ภาพที่ 1.1 เดนโดรแกรมที่ได้จากการวิเคราะห์ข้อมูลด้วยการวิเคราะห์จัดกลุ่ม (จาก Brown, 1991)



ภาพที่ 1.2 อิซิเคิลที่ได้จากการวิเคราะห์ข้อมูลด้วยการวิเคราะห์จัดกลุ่ม (จาก Sneath, 1957)

ตัวอย่างการศึกษาโดยใช้เทคนิคนี้พอกล่าวได้คือ การศึกษาของ Macfie, Gutteridge and Norris (1978) ในแบคทีเรีย 5 กลุ่มรวม 25 สายพันธุ์ (strain) ที่มีความใกล้ชิดกัน การวิเคราะห์ข้อมูลใช้เทคนิค 3 เทคนิคคือ การวิเคราะห์ปัจจัย การวิเคราะห์จัดกลุ่ม และ canonical variate analysis โดยใช้ข้อมูลความสูงของกราฟที่ได้จากการทำโครมาโตรกราฟ 24 ยอด (peak) ผลการวิเคราะห์ปัจจัย พบว่าการกระจายของ factor score ของแต่ละสายพันธุ์ไม่มีการเกาะกลุ่มกันเลย ผลการวิเคราะห์จัดกลุ่มพบว่าความเหมือนกันของสายพันธุ์ระหว่างกลุ่มและภายในกลุ่มมีความแตกต่างกันน้อยมาก ส่วนผลการวิเคราะห์ด้วย canonical variate analysis พบว่าการกระจายของแต่ละสายพันธุ์บนแกน canonical variate ที่ 1 และ 2 ที่เป็นกลุ่มเดียวกันจะเกาะกลุ่มกันและแยกจากกลุ่มอื่นค่อนข้างเด่นชัด โดยเฉพาะกลุ่มที่ 4 แยกออกจากอีก 4 กลุ่มที่เหลืออย่างเห็นได้ชัด ส่วนกลุ่มที่ 1 2 3 และ 5 มีความใกล้ชิดกันมาก

การนำเอาเทคนิคการวิเคราะห์ PCA, Principal Coordinate Analysis, Discriminant Analysis และ Cluster Analysis มาใช้ในการศึกษา numerical taxonomy ของพืชพอกล่าวได้โดยสรุปดังนี้

Reynold and Crawford (1980) ศึกษา variation ของพืชชนิดซับซ้อน (species complex) Chenopodium atrovirens-desiccatum-pratericola ด้วยการวิเคราะห์จัดกลุ่ม โดยใช้ลักษณะของพืช 14 ลักษณะพบว่าพืชชนิดซับซ้อนนี้ประกอบด้วย 3 กลุ่ม ซึ่งเมื่อกำหนดชนิดของแต่ละ OTU ที่นำมาศึกษาพบว่าแต่ละกลุ่มคือ C. atrovirens , C. pratericola และ C. desiccataum การวิเคราะห์ด้วยเทคนิคการวิเคราะห์ปัจจัย พบว่า factor score บนแกน factor ที่ 1 และ 2 มีการกระจายแยกเป็น 3 กลุ่ม และเมื่อระบุชนิดของแต่ละ OTU พบว่า ทั้งสามกลุ่มมีสมาชิกเป็นแต่ละชนิดดังผลการวิเคราะห์จัดกลุ่มและการวิเคราะห์ด้วยเทคนิค canonical variate analysis เพื่อตรวจสอบการแยกของแต่ละกลุ่มโดยระบุความเป็นสมาชิกของแต่ละ OTU จากชนิดของตัวอย่างพบว่าการกระจายของค่าคะแนนจัดจำแนกบนแกนที่ 1 และ 2 ซึ่งมี 3 กลุ่ม แต่ละกลุ่มมีการเหลื่อมล้ำกันบ้าง และมีเปอร์เซ็นต์การทำนายกลุ่มจากสมการสูงถึง 91 เปอร์เซ็นต์ขึ้นไป สำหรับตัวอย่างที่ยังไม่สามารถระบุได้ว่าเป็นชนิดใดแน่ สามารถหาชื่อชนิดได้ในการวิเคราะห์ขั้นนี้ อย่างไรก็ตามจากผลการวิเคราะห์ลักษณะของพืชชนิดซับซ้อนนี้ พบว่าไม่สามารถใช้ลักษณะเพียงลักษณะเดียวแยกพืชทั้ง 3 ชนิดออกจากกันได้ แต่ต้องใช้ลักษณะมากกว่า 2 ลักษณะประกอบกับการกระจายพันธุ์ของแต่ละชนิดจึงจะสามารถระบุชื่อชนิดได้

Small (1987) ศึกษาพืช 23 ชนิดในสกุล Trigonella ที่มีสารประกอบในต้นคล้ายกับพืชในสกุล Medicago ซึ่งอยู่ในวงศ์เดียวกัน เดิมมีการจัดกลุ่มภายในเป็น 3 sect. 4 subsect. และ 5 series ดังนี้

"medicagoid" Trigonella spp.
- sect. Bucerates Boiss.
  - subsect. Reflexae Sirj.
  - subsect. Erectae Sirj.
    - ser. Astroites Sirj.
    - ser. Curvatae Sirj.
    - ser. Crassipedes Sirj.
    - ser. Sessiles Sirj.
    - ser. Reflexae Sirj.
- sect. Isthmocarpae Boiss.
- sect. Lunatae Boiss
  - subsect. Simplices Sirj.
  - subsect. Glanduliferae Sirj.

ซึ่งการจัดกลุ่มดังกล่าวไม่มีความแน่ชัดของลักษณะที่ใช้ในการแบ่งกลุ่ม ในการตรวจสอบสถานะของหน่วยทางอนุกรมวิธานดังกล่าวภายในพืชกลุ่มนี้ทำโดยวิเคราะห์ลักษณะของพืชทั้ง 23 ชนิดจำนวน 46 ลักษณะด้วยการวิเคราะห์จัดกลุ่ม พบว่า พืชกลุ่มนี้ประกอบด้วย 2 กลุ่มย่อย (I และ II) โดยที่กลุ่มย่อย I ภายในมีเพียง 2 กลุ่มย่อยเท่านั้น (a และ b) สำหรับกลุ่มย่อย II ประกอบด้วย 4 กลุ่มย่อย (A B C และ D) โดยในกลุ่ม A มี 3 กลุ่มย่อย (1 2 3) และการวิเคราะห์ด้วยเทคนิค Principal Coordinate Analysis (PCoA) พบว่าพืชชนิดนี้ประกอบด้วย 2 กลุ่มโดยกลุ่มย่อยภายในมีสมาชิกเหมือนกับการวิเคราะห์ด้วยการวิเคราะห์จัดกลุ่ม เมื่อกำหนดชนิดของแต่ละ OTU ลงในเดนโดรแกรมและค่าคะแนนที่ได้จาก PCoA พบว่าแต่ละกลุ่มย่อยมีหน่วยทางอนุกรมวิธานที่เป็นสมาชิกดังนี้

กลุ่มย่อย I ประกอบด้วย sect. Lunatae
- กลุ่ม a มี subsect. Simplices
- กลุ่ม b มี subsect. Glanduliferae

กลุ่มย่อย II ประกอบด้วยพืชที่เหลือ
- กลุ่ม I-A-1 ประกอบด้วย 1 series ใน sect. Bucerates, subsect. Erectae คือ ser. Astroites
- กลุ่ม I-A-2 ประกอบด้วย 3 series ใน sect. Bucerates, subsect. Erectae คือ ser. Sessiles ser. Curvatae

และ ser. Crassipedes

กลุ่ม I-A-3 ประกอบด้วยพืชที่ยังไม่สามารถจัดได้ว่าควรอยู่ในกลุ่มใด

กลุ่ม I-B ประกอบด้วยพืชที่อยู่ใน sect. Bucerates subsect. Erectae ser. Reflexae

กลุ่ม I-C ประกอบด้วยพืชที่อยู่ใน sect. Bucerates, subsect. Reflexae

กลุ่ม I-D ประกอบด้วยพืชที่อยู่ใน sect. Isthmocarpae

Doebley and Iltis (1980) ได้ศึกษาพืชในสกุล Zea 4 ชนิด คือ Z. mays Linn. (s.l.) (ซึ่งมี 3 subspecies คือ Z. mays ssp. mays, Z. mays ssp. mexicana (Schrader) Iltis และ Z. mays ssp. parviglumis Iltis and Doebley) Z. luxurians (Durien and Ascherson) Bird Z. perennis (Hicthc. Reeves and Mangelsdorf และ Z. diplopereunnis Iltis ซึ่งเดิมแบ่งพืชทั้ง 4 ชนิดออกเป็น 2 section คือ section Zea ประกอบด้วย Z. mays ssp. mays เพียง subspecies เดียว และ section Euchlaena ซึ่งประกอบด้วย พืชที่เหลืออีก 3 ชนิดกับ 2 subspecies ในการแบ่งเป็น section ดังกล่าวใช้ลักษณะของดอกตัวเมียเป็นเกณฑ์ ซึ่ง Doebley และ Iltis กล่าวว่ามีการเปลี่ยนได้ง่ายถ้ามีการปรับปรุงพันธุ์ ดังนั้นจึงใช้ลักษณะของดอกตัวผู้แทน โดยวัดลักษณะ 10 ลักษณะมาทำการวิเคราะห์ด้วยเทคนิค discriminant analysis พบว่าการกระจายของคะแนนจัดจำแนกบนแกนสมการที่ 1 และ 2 แบ่งพืชออกเป็น 2 กลุ่ม และมี rank เป็น section คือ กลุ่ม I ประกอบด้วย Z. mays ทั้ง 3 subspcies และกลุ่ม II ประกอบด้วย Z. luxurians Z. diploperennis และ Z. perennis และได้กำหนดชื่อสำหรับกลุ่ม I คือ sect. Zea และกลุ่ม II คือ sect. Luxuriantes

Standley (1987) ได้ทำการศึกษาพืชชนิดซับซ้อน Carex lenticularis Michx. ในภาคตะวันออกของทวีปอเมริกาเหนือ 2 ชนิด คือ C. lenticularis var.lenticularis และ C. nigra (Linn.) Reich. ซึ่งเดิมเข้าใจว่าในพืชชนิดแรกมี 5 วาไรตี (variety) คือ C. lenticularis var. lenticularis, C. lenticularis var. albimontana Dewey, C. lenticularis var. Blakei Dewey, C. lenticularis var.merens Howe และ C. lenticularis var. eucycla Fernald & Wiegand ส่วนพืชชนิดหลังประกอบด้วย 2 วาไรตี คือ C. nigra var. nigra และ C. nigra var. strictiformis (Bailey) Fernald ในการศึกษาเพื่อทำการตรวจสอบสถานะของหน่วยทางอนุกรมวิธานภายในพืชทั้ง 2 ชนิดดังกล่าวใช้ลักษณะ 44 ลักษณะจาก 89 ตัวอย่างที่เป็น C.

nigra และ 101 ตัวอย่างที่เป็น C. lenticularis var. lenticularis นำมาวิเคราะห์จัดกลุ่มและ PCA โดยวิเคราะห์แยกชนิดกัน ผลการวิเคราะห์ใน C. lenticularis var. lenticularis ด้วยการวิเคราะห์จัดกลุ่ม พบว่า ไม่มีกลุ่มย่อยในพืชชนิดนี้ ประกอบกับผลการวิเคราะห์ด้วย PCA ที่ factor score บนแกน factor ที่ 1 และ 2 ไม่มีการแยกกลุ่มกัน แสดงว่า C. lenticularis ที่กระจายพันธุ์อยู่ในภาคตะวันออกทวีปอเมริกาเหนือมีเพียง 1 วาไรตี คือ C. lenticularis var.lenticularis สำหรับผลของการวิเคราะห์จัดกลุ่มและ PCA ใน C. nigra พบว่า ไม่มีหน่วยทางอนุกรมวิธานที่มี rank ต่ำกว่าชนิด (infraspecific taxon) ในพืชชนิดนี้ นอกจากนี้ยังได้ทำการศึกษา C. lenticularis ที่กระจายพันธุ์อยู่บริเวณตอนกลางไปจนถึงภาคตะวันตกซึ่งมี 5 วาไรตีเปรียบเทียบกับพืชที่กระจายพันธุ์อยู่ในภาคตะวันออกด้วยเทคนิค PCA พบว่า factor score บนแกน factor ที่ 1 และ 2 มีการกระจายแยกเป็น 2 กลุ่ม เมื่อกำหนดพื้นที่การกระจายพันธุ์ให้แต่ละ OTU พบว่ากลุ่มที่ 1 เป็นพืชที่มีการกระจายพันธุ์อยู่ในภาคตะวันออก แต่กลุ่มที่ 2 เป็นพืชที่มีการกระจายพันธุ์อยู่ในตอนกลางและภาคตะวันตก แต่เมื่อกำหนดชนิดให้กับแต่ละ OTU พบว่า กลุ่มที่ 1 คือ C. lenticularis var. lenticularis และในกลุ่มที่ 2 ประกอบด้วยพืชอีก 5 วาไรตีที่กระจายพันธุ์อยู่บริเวณตอนกลางไปจนถึงภาคตะวันตกและยังพบว่าแต่ละวาไรตีไม่มีการแยกกันเป็นแต่ละกลุ่ม อย่างไรก็ตาม Standley ไม่ได้ทำการศึกษาสถานะของพืชที่มีการกระจายพันธุ์ในตอนกลางและภาคตะวันตกจึงไม่ได้กล่าวถึงพืชในพื้นที่ดังกล่าวอย่างละเอียด จากที่กล่าวมาทั้งหมดสรุปได้ว่า C. lenticularis ที่มีการกระจายพันธุ์อยู่ในภาคตะวันออกของทวีปอเมริกาเหนือมีเพียง 1 หน่วยทางอนุกรมวิธานที่มีอยู่จริงและมี rank เป็น วาไรตี คือ C. lenticularis Michx. var. lenticularis ส่วน C. nigra เป็นหน่วยทางอนุกรมวิธานที่ยังไม่มีหลักฐานที่จะแยกออกเป็นพืชในระดับที่ต่ำกว่าชนิด

Heard and Semple (1988) ทำการตรวจสอบสถานะของหน่วยทางอนุกรมวิธานภายในพืชชนิดซับซ้อน Solidago rigida Linn. ซึ่งมีการกระจายพันธุ์อยู่ตั้งแต่ภาคตะวันออกจนถึงภาคกลางตอนเหนือของประเทศสหรัฐอเมริกา ภาคกลางตอนใต้และภาคตะวันตกของประเทศแคนาดา จากการที่มีพื้นที่ในการกระจายพันธุ์กว้างและแต่ละพื้นที่จะพบต้นพืชที่มีลักษณะต่าง ๆ กันไป จึงมีการแยกออกเป็นชนิดใหม่และตั้งชื่อขึ้นทั้งในระดับที่เป็น species subspecies และ variety จนทำให้เกิดความสับสนในพืชชนิดซับซ้อนนี้ อย่างไรก็ตามสามารถแบ่งพืชตามการกระจายพันธุ์และลักษณะสัณฐานวิทยาออกได้เป็น 3 กลุ่ม ดังนั้น Heard and Semple จึงมุ่งตรวจสอบสถานะของพืชทั้ง 3 กลุ่มว่ามีอยู่จริงหรือไม่ วิธีการที่นำมาใช้คือการวิเคราะห์ลักษณะสัณฐานวิทยา 18 ลักษณะจากตัวอย่างพืชทั้ง 3 กลุ่มรวม 103 OTU ด้วยการวิเคราะห์จัดกลุ่ม พบว่าแบ่งพืชออกได้เป็น 2 กลุ่ม (A และ B) โดยที่ ในกลุ่ม A ประกอบด้วยพืชที่มีinvolucral scale ไม่มีขนปกคลุม ส่วนกลุ่ม B ประกอบด้วยพืชที่มีขนปกคลุมที่ involucral scale และ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภายในกลุ่ม B ยังประกอบด้วยกลุ่มย่อย 2 กลุ่ม (1 และ 2) ซึ่งกลุ่มย่อย 1 มีพืชที่มีขนที่ involucral scale ปกคลุมหนาแน่นเป็นสมาชิกกลุ่ม แต่กลุ่มย่อย 2 มีพืชที่มีขนประปรายที่ involucral scale จากนั้นนำข้อมูลชุดเดียวกันมาวิเคราะห์ด้วยเทคนิคการวิเคราะห์จัดจำแนก โดยใช้ผลของการวิเคราะห์จัดกลุ่มมากำหนดกลุ่ม พบว่า ความถูกต้องของการทำนายกลุ่มจากสมการมีค่ามากกว่า 95 เปอร์เซ็นต์ และจากการกระจายของค่าคะแนนจัดจำแนกบนแกนสมการที่ 1 และ 2 พบว่า กลุ่ม A และกลุ่ม B1 แยกกันอย่างเด่นชัด แต่กลุ่ม A กับกลุ่ม B2 และกลุ่ม B1 กับกลุ่ม B2 มีการการจายของคะแนนซ้อนทับกันเล็กน้อย จากผลการวิเคราะห์แสดงว่าพืชทั้งสามกลุ่มแยกกันเป็นแต่ละ taxon แล้ว แต่ยังมีความใกล้ชิดกันอยู่ เพราะจากการกระจายค่าคะแนนจัดจำแนกที่แต่ละกลุ่มอยู่ใกล้ ๆ กัน และเมื่อนำการกระจายพันธุ์มาพิจารณาประกอบกัน ทำให้ตัดสินว่าพืชทั้ง 3 กลุ่มยังคงอยู่ใน species เดียวกัน แต่มี rank เป็น subspecies โดยที่กลุ่ม A ซึ่งมีการกระจายพันธุ์ทางภาคตะวันออกเฉียงใต้ของประเทศสหรัฐอเมริกา คือ S. rigida Linn. ssp. glabrata (Braun) Heard & Semple กลุ่ม B1 ซึ่งมีการกระจายพันธุ์อยู่ตั้งแต่ภาคกลางตอนเหนือของประเทศสหรัฐอเมริกาจนถึงภาคกลางตอนใต้และภาคตะวันตกของประเทศแคนาดา คือ S. rigida Linn. ssp. humilis (Poster) Herad & Semple และกลุ่ม B2 ซึ่งมีการกระจายพันธุ์ตั้งแต่ภาคกลางตอนเหนือจนถึงภาคตะวันออกเฉียงเหนือของประเทศสหรัฐอเมริกา คือ S. rigida Linn. ssp. rigida

Menadue and Growden (1988) ศึกษาตรวจสอบสถานะของ Ranunculus decurvus (Hook f.) Melville และ R. concinnus (Hook f.) Melville ว่าเป็นพืชชนิดเดียวกันหรือเป็นแต่ละชนิด ทั้งนี้เนื่องจากรูปวิธานของพืชทั้งสองไม่สามารถใช้จำแนกตัวอย่างในธรรมชาติได้ว่าเป็นชนิดใด ในการตรวจสอบได้ใช้ลักษณะสัณฐานวิทยาทั้งสิ้น 43 ลักษณะ โดยแบ่งข้อมูลเป็น 3 ชุดคือ ชุดที่ 1 ประกอบด้วยลักษณะที่ใช้จำแนกชนิดของพืชทั้งสองชนิด ชุดที่ 2 ประกอบด้วยลักษณะทุกลักษณะ และชุดที่ 3 ประกอบด้วยลักษณะที่มีการปรับขนาด โดยการวิเคราะห์ด้วยเทคนิค PCA และการวิเคราะห์จัดจำแนก พบว่า factor score ของข้อมูลชุดที่ 1 บนแกน factor ที่ 1 และ 2 มีการกระจายเป็นกลุ่มเพียง 1 กลุ่ม แม้ว่าเมื่อดูแกน factor ที่ 2 จะมีการแยกกันของ factor score บ้าง แต่ไม่เด่นชัดนัก และผลการวิเคราะห์จัดจำแนกของข้อมูลชุดเดียวกันพบว่าได้ผลว่าฮิสโตแกรมของค่าคะแนนจัดจำแนกของพืชทั้งสองชนิดซ้อนทับกันมาก สำหรับการวิเคราะห์ข้อมูลชุดที่ 2 และ 3 ด้วยเทคนิคเดียวกับข้างต้น ให้ผลลัพธ์เดียวกันกับผลลัพธ์ข้างต้น และการวิเคราะห์โดยใช้เทคนิคการวิเคราะห์จัดกลุ่มของลักษณะทั้งหมด โดยรวมตัวอย่างต้นแบบ (type specimen) ของพืชทั้งสองชนิดเข้าในการวิเคราะห์ด้วย พบว่า ทุกค่าสังเกตรวมกันเป็น 1 กลุ่มเท่านั้น นั่นคือ Ranunculus decurvus และ R. concinnus เป็นพืชชนิดเดียวกัน สำหรับชื่อที่ใช้เรียกพืชภายหลังการรวมพืชทั้งสองชนิดเป็นชนิดเดียวกันนั้น เนื่องจากได้มีการจัดทำคำบรรยายลักษณะของ Ranunculus decurvus (Hook f.) Melville

าอน ซึ่งตามกฎของการตั้งชื่อแล้ว Ranuculus decurvus (Hook f.) Melville เป็นชื่อที่มี priority สูงกว่า R. concinnus (Hook f.) Melville ดังนั้นจึงใช้ชื่อ Ranuculus decurvus (Hook f.) Melville เรียกพืชที่ภายหลังจากรวม Ranunculus decurvus และ R. concinnus เป็นพืชชนิดเดียวกัน

Downie and McNeill (1990) ทำการตรวจสอบสถานะของ taxa ภายในพืชชนิดซับซ้อน Euphrasia randii Reeks โดยการวิเคราะห์ลักษณะ 13 ลักษณะในพืช 291 ต้นจาก 59 ประชากรด้วยเทคนิค PCA และการวิเคราะห์จัดกลุ่ม ผลการวิเคราะห์ด้วย PCA พบว่าการกระจายของ factor score บนแกน factor ที่ 1 และ 2 ไม่มีการแยกเป็นกลุ่มแต่กระจายปะปนกัน และจากเดนโดรแกรมที่ได้จากการวิเคราะห์จัดกลุ่มไม่สามารถกำหนดกลุ่มย่อยภายในพืชชนิดซับซ้อนนี้ ดังนั้น Downei and McNeill จึงสรุปว่าพืชชนิดซับซ้อนนี้ประกอบด้วยพืชเพียง 1 ชนิดเท่านั้น คือ E. randii

Semple, Chmielewski and Brammal (1990) ได้ศึกษาพืชชนิดซับซ้อน Solidago nemoralis Aiton ซึ่งมีการกระจายพันธุ์อยู่บริเวณตอนกลางและภาคตะวันออกของทวีปอเมริกาเหนือ พืชที่มีการกระจายอยู่ภาคตะวันออกมีจำนวนโครโมโซม 2n=18 และ 36 แต่พืชที่กระจายอยู่ตอนกลางมีจำนวนโครโมโซม 2n=36 การศึกษาทางด้าน numerical taxonomy ของพืชชนิดซับซ้อนนี้ใช้ลักษณะ 11 ลักษณะใน 177 OTU มาวิเคราะห์ด้วยเทคนิคการวิเคราะห์จัดกลุ่ม พบว่า เดนโดรแกรมที่ได้ประกอบด้วย 2 กลุ่ม (A และ B) โดยที่กลุ่ม A ประกอบด้วยพืชที่มีจำนวนโครโมโซม 2n=36 และมีการกระจายพันธุ์อยู่บริเวณด้านตะวันตกของการกระจายพันธุ์ ส่วนกลุ่ม B เป็นพืชที่มีการกระจายพันธุ์อยู่บริเวณตอนกลางและด้านตะวันออกของการกระจายพันธุ์ ซึ่งในกลุ่มนี้ประกอบด้วย 2 กลุ่มย่อยคือ กลุ่มย่อย I ประกอบด้วยพืชที่มีจำนวนโครโมโซม 2n=18 และ กลุ่มย่อย II ประกอบด้วยพืชที่มีจำนวนโครโมโซม 2n=36 จากนั้นนำข้อมูลชุดเดียวกันไปวิเคราะห์ด้วยเทคนิคการวิเคราะห์จัดจำแนกโดยใช้ผลจากการวิเคราะห์จัดกลุ่มมากำหนดกลุ่มของ OTU ซึ่งผลการวิเคราะห์พบว่าการกระจายของค่าคะแนนจัดจำแนกบนแกนสมการที่ 1 และ 2 ของทั้ง 3 กลุ่มมีการแยกออกจากกันค่อนข้างมากโดยเฉพาะกลุ่ม A แยกจากกลุ่ม BI อย่างเด่นชัด แต่กลุ่ม BI กับกลุ่ม BII และกลุ่ม A กับกลุ่ม BII มีการกระจายซ้อนทับกันบ้าง ความถูกต้องของการทำนายกลุ่มจากสมการมีค่า 88 84 และ 93 เปอร์เซ็นต์ในกลุ่ม A BI และ BII ตามลำดับ สำหรับการกำหนดชนิดภายในพืชชนิดซับซ้อนนี้ Semple et al. ได้กล่าวว่ากลุ่ม BI และกลุ่ม BII ไม่น่าจะแยกออกจากกันเป็นแต่ละชนิดที่แตกต่างกันเพราะลักษณะที่แยกพืชทั้งสองกลุ่มออกจากกันเป็นลักษณะทางปริมาณและเป็นผลเนื่องมาจากจำนวนโครโมโซม ดังนั้นพืชชนิดซับซ้อน S. nemoralis จึงประกอบด้วย 2 ชนิดและมี rank เป็น subspecies เพราะมีการกระจายพันธุ์ต่างพื้นที่กัน ทั้ง 2 subspecies คือ S. nemoralis

Aiton ssp. nemoralis กระจายพันธุ์ทางตะวันออกของบริเวณที่มีการกระจายพันธุ์ และ S. nemoralis Aiton ssp. decemflora (DC.) Brammall กระจายพันธุ์ทางตะวันตกของบริเวณที่มีการกระจายพันธุ์

Small and Brooks (1991) ทำการตรวจสอบสถานะของ Medicago sinkiae Uljanova ซึ่งได้จัดไว้ใน sect. Spirocarpos Ser. subsect. Pachyspirae (Urb.) Heyn พืชชนิดนี้ยังเป็นที่สงสัยกันว่าเป็นพืชที่มีอยู่จริงหรือไม่ การตรวจสอบได้ใช้ลักษณะ 30 ลักษณะจากพืชอีก 12 ชนิดที่อยู่ใน subsect. เดียวกัน แล้วนำข้อมูลมาวิเคราะห์จัดกลุ่ม จากเดนโดรแกรมที่ได้ พบว่า พืชใน subsect. นี้แบ่งเป็น 5 กลุ่มซึ่งแต่ละกลุ่มไม่ค่อยมีความใกล้ชิดกันมาก และสำหรับ M. sinskiae นั้นเป็นพืชที่มีอยู่จริงและมีความใกล้ชิดกับกลุ่ม M. rigida-M. rigiduloides และจากผลของ principal coordinate analysis ของข้อมูลชุดเดียวกันพบว่า M. sinskiae เป็นพืชที่มีอยู่จริงและไม่มีความใกล้ชิดกับชนิดอื่น ๆ เลย แม้ว่าจะอยู่ใกล้กับกลุ่ม M. rigida-M. rigiduloides แต่ก็นับว่าอยู่ห่างกันเมื่อดูจากการกระจายของคะแนนบน principal coordinate axis I และ II ดังนั้น M. sinskiae Uljanova จึงเป็นชนิดที่มีอยู่จริง

จากที่กล่าวมาข้างต้นลักษณะที่นำมาใช้ในการศึกษา numerical taxonomy ได้แก่ ลักษณะของต้นและใบ และลักษณะดอกเป็นหลัก อย่างไรก็ตามได้มีการนำเอาลักษณะที่มีขนาดเล็ก เช่น ละอองเรณู เป็นต้น มาเป็นลักษณะที่ใช้ในการศึกษา numerical taxonomy ดังจะกล่าวเป็นตัวอย่างได้ดังนี้

Hanks and Fryxell (1979) ศึกษาลักษณะสัณฐานวิทยาของละอองเรณูของพืชในสกุล Herissantia 4 ชนิด และพืชสกุล Gaya 6 ชนิด ซึ่งมีปัญหาเกิดขึ้นเนื่องจากว่า 1)Gaya elingulata มีจำนวนโครโมโซม 2n=12 ซึ่งเหมือนกันกับพืชในสกุลเดียวกันแต่ไม่มีโครงสร้างที่เรียกว่า endoglassum ที่เมล็ด และ 2) Herissantia tiubae มีลักษณะสัณฐานวิทยาเหมือนพืชในสกุลเดียวกัน แต่มีจำนวนโครโมโซม 2n=12 ซึ่งต่างจากพืชในสกุลเดียวกันที่มีจำนวนโครโมโซม 2n=14 และจากปัญหาข้างต้นทำให้เกิดความคิดว่า 1) H. tiubae คงอยู่ในสกุล Herissantia และสรุปว่าพืชในสกุลนี้มีจำนวนโครโมโซมหลายแบบ หรือ 2) ควรย้าย H. tiubae ไปไว้ในสกุล Gaya เพื่อให้พืชทั้งสองชนิดมีจำนวนโครโมโซมเพียงแบบเดียวในแต่ละสกุล หรือ 3) รวมพืชทั้งสองสกุลเข้าด้วยกัน Hanks and Fryxell ได้ศึกษาลักษณะของละอองเรณู 9 ลักษณะของพืชทั้ง 2 สกุล (G. elingulata ไม่ได้นำมาศึกษา) นำมาวิเคราะห์ด้วยการวิเคราะห์จัดกลุ่ม ได้ผลว่า พืชทั้ง 10 ชนิดแบ่งเป็น 2 กลุ่ม คือ กลุ่มที่ประกอบด้วยพืชในสกุล Gaya และกลุ่มที่ประกอบด้วยพืชในสกุล Herissantia จากเดนโดรแกรมและกราฟ

ระหว่างลักษณะ 2 ลักษณะจำนวน 4 คู่ที่เลือกมาได้แสดงให้เห็นว่า H. tiubae มีความใกล้ชิดกับพืชในสกุลเดียวกันมากกว่า ดังนั้นจึงสรุปได้ว่าลักษณะสัณฐานวิทยาของละอองเรณูของ H. tiubae เหมือนกับละอองเรณูของพืชสกุล Herissantia มากกว่าจะเหมือนกับละอองเรณูของพืชในสกุล Gaya นั้นคือ H. tiubae ควรอยู่ในสกุลเดิม

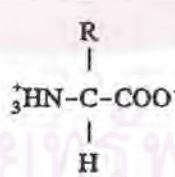
Robbins, Dickinson, and Rhodes (1979) ได้ศึกษาละอองเรณูของพืชในสกุล Ambrosia 4 ชนิด คือ A. trifida, A. bidenta, A. artemisiifolia และ A. psilostachya โดยใช้ลักษณะ 6 ลักษณะซึ่งวัดจากภาพถ่ายจากกล้องจุลทรรศน์อิเลกตรอน แล้วนำมาวิเคราะห์ด้วยเทคนิคการวิเคราะห์จัดจำแนกซึ่งแบ่งเป็น 3 ตอน คือ ตอนที่ 1 ใช้ละอองเรณูของพืชทั้ง 4 ชนิดมาวิเคราะห์ได้ผลว่าสามารถทำนายกลุ่มประชากรของละอองเรณูจากสมการได้ถูกต้องร้อยละ 86 และเมื่อนำสมการดังกล่าวมาใช้ทำนายกลุ่มประชากรของละอองเรณูที่ได้จากตัวอย่างแห้งพบว่าสามารถทำนายได้ถูกต้องร้อยละ 100 58 100 และ 100 สำหรับละอองเรณูของ A. trifida, A. artemisiifolia, A. bidenta และ A. psilostachya ตามลำดับ ตอนที่ 2 เมื่อตัดละอองเรณูของ A. psilostachya ออกแล้ววิเคราะห์จัดจำแนกละอองเรณูของพืชทั้ง 3 ชนิดที่เหลือ พบว่า สามารถทำนายกลุ่มประชากรของละอองเรณูจากสมการได้ถูกต้องร้อยละ 88 88 และ 84 สำหรับละอองเรณูของ A. trifida, A. artemisiifolia และ A. bidenta ตามลำดับ และเมื่อนำสมการมาใช้ทำนายกลุ่มประชากรของละอองเรณูจากตัวอย่างแห้งพบว่าสามารถใช้ทำนายกลุ่มประชากรของละอองเรณูได้ถูกต้องร้อยละ 100 75 และ 100 ตามลำดับ ตอนที่ 3 หลังจากตัดละอองเรณูของ A. bidenta และ A. psilostachya แล้ววิเคราะห์จัดจำแนกละอองเรณูของพืชอีก 2 ชนิดที่เหลือได้ผลว่าสามารถใช้สมการจัดจำแนกทำนายกลุ่มประชากรของละอองเรณูได้ถูกต้องร้อยละ 98 และ 97 สำหรับ A. trifida และ A. bidenta ตามลำดับ และเมื่อนำสมการดังกล่าวมาใช้ทำนายกลุ่มของละอองเรณูที่ได้จากตัวอย่างแห้ง พบว่า สามารถทำนายกลุ่มประชากรของละอองเรณูได้ถูกต้องร้อยละ 100 และ 92 ตามลำดับ จากการศึกษาทั้งหมด Robbins et al. จึงสรุปว่าสามารถใช้ลักษณะสัณฐานวิทยาของละอองเรณูมาตรวจหาชนิดพืชในสกุลนี้ได้ และเป็นประโยชน์ในการวินิจฉัยสาเหตุของผู้ป่วยที่เป็นโรคหอบเนื่องจากละอองเรณูพืช อีกทั้งสามารถใช้ในการแยกฟอสซิลพืชในสกุลนี้โดยเปรียบเทียบลักษณะละอองเรณู

Isozyme Electrophoresis

ไอโซไซม์ (isozyme) คือ เอนไซม์ที่มีโครงสร้างโมเลกุลหลายแบบ แต่ละแบบมีความจำเพาะต่อปฏิกิริยาเดียวกัน (Markert and Moller, 1959) คำที่มีความหมายคล้ายกันแต่ค่อนข้างเฉพาะเจาะจงและนักชีวเคมีใช้กันมากคือ ไอโซเอนไซม์ (isoenzyme หรือ iso-enzyme) ซึ่งมีความหมายว่า "เอนไซม์ในสิ่งมีชีวิต 1 ชนิดที่มีโครงสร้างโมเลกุลหลายแบบ แต่ละแบบมีความจำเพาะต่อปฏิกิริยาเดียวกัน" (Webb, 1964) อย่างไรก็ตามในปัจจุบันสามารถใช้คำทั้งสองแทนกันได้ ไอโซไซม์แต่ละแบบอาจสร้างจากจีน (gene) ต่างตำแหน่งหรือสร้างจากตำแหน่งเดียวกันแต่ต่างอัลลีล (allele) กันก็ได้ (Nelson and Burr, 1973) ซึ่งจะเรียกเอนไซม์ที่สร้างมาจากจีนที่ตำแหน่งเดียวกันแต่ต่างอัลลีลกันว่า **"อัลโลไซม์"** (allozyme) (Prakash, Lewontin and Hubby, 1969) ในการศึกษาไอโซไซม์จะใช้วิธีการทางชีวเคมีอย่างใดอย่างหนึ่ง ซึ่งได้แก่ 1) electrophoresis 2) chromatography 3) gel filtration 4) immunochemistry และ 5) sedimentation แต่วิธีที่ให้ผลดีที่สุดคือ electrophoresis (Shannon, 1968)

หลักการของ eletrophoresis มีพื้นฐานพอกล่าวโดยย่อได้ดังนี้ (Murphy et al., 1990)

โปรตีนคือสายโพลีเปปไทด์ที่มีกรดอะมิโนมาต่อกันด้วยพันธะเปปไทด์ โดยลำดับของกรดอะมิโนของสายโพลีเปปไทด์ถูกควบคุมโดยรหัสพันธุกรรม กรดอะมิโนมีสูตรโครงสร้างดังนี้



โดยหมู่ -COOH จะปล่อย H^+ ออกมาและ $-NH_2$ จะรับ H^+ กลายเป็น $-COO^-$ และ $-NH_3^+$ ตามลำดับ หมู่ R อาจเป็นไฮโดรเจน เช่น กรดอะมิโนไกลซีน (glycine) เป็นต้นหรืออาจเป็นไฮโดรคาร์บอน เช่น กรดอะมิโนซีรีน (serine) เป็นต้น นอกจากนี้อาจมีหมู่ $-NH_2$ มาต่อกับหมู่ R ทำให้กรดอะมิโนมีสมบัติเป็นเบสและมีประจุสุทธิเป็นบวก (เนื่องจาก $-NH_2$ รับ H^+ กลายเป็น $-HN_3^+$) เช่น กรดอะมิโนไลซีน (lysine) กรดอะมิโนอาจีนีน (arginine) เป็นต้น แต่ถ้ามีหมู่ -COOH มาต่อกับหมู่ R ทำให้กรดอะมิโนมีสมบัติเป็นกรดและมีประจุสุทธิเป็นลบ (เนื่องจาก -COOH ปล่อย H^+ กลายเป็น $-COO^-$) ดังนั้นประจุสุทธิของโปรตีนจึงขึ้นอยู่กับผลรวมของประจุสุทธิของกรดอะมิโนที่เป็นองค์ประกอบ นอกจากนี้ pH ของสารละลายยังมีผลต่อประจุของกรดอะมิโนด้วย กล่าวคือ ถ้าสารละลายมี pH ต่ำ กรดอะมิโนจะมีประจุบวก (เนื่องจากที่ pH ต่ำ มี

H^+ มาก หมู่ -COOH จึงไม่ปล่อย H^+ ออกมา แต่ $-NH_2$ จะรับ H^+ จากสารละลายเป็น $-NH_3^+$) แต่ถ้าสารละลายมี pH สูง กรดอะมิโนจะมีประจุลบ (เนื่องจากที่ pH สูง มี OH^- มาก หมู่ -COOH จะปล่อย H^+ ออกมาและกลายเป็น $-COO^-$ และหมู่ $-NH_2$ ไม่รับ H^+ เพราะในสารละลายมี H^+ น้อย และถ้าหมู่ $-NH_2$ อยู่ในสภาพ $-NH_3^+$ ก็จะปล่อย H^+ ออกมาและกลายเป็น $-NH_2$ ในที่สุด) นั่นคือประจุสุทธิของโปรตีนและ pH ของสารละลายจะมีผลต่อการเคลื่อนที่ของโปรตีนในสนามไฟฟ้า

โครงสร้างของโมเลกุลโปรตีนอาจแบ่งได้ 4 โครงสร้างคือ 1) โครงสร้างขั้นที่ 1 เป็นโครงสร้างที่ได้ภายหลังเกิดการแปลรหัสพันธุกรรม 2) โครงสร้างขั้นที่ 2 เกิดจากโมเลกุลโปรตีนที่มีโครงสร้างขั้นที่ 1 มีการพับ (folding) หรือพันเป็นบันไดเวียน (helix) ทั้งนี้อาจมีพันธะซัลไฟด์ได้ด้วย 3) โครงสร้างขั้นที่ 3 เกิดจากโมเลกุลโปรตีนที่มีโครงสร้างขั้นที่ 2 มีการพับ (folding) และ 4) โครงสร้างขั้นที่ 4 เกิดสายโพลีเปปไทด์มากกว่า 1 สายมาประกอบกันซึ่งอาจจับกันด้วยพันธะไฮโดรเจน พันธะไอออนิก หรือแรงแวนเดอวาล โครงสร้างโมเลกุลแต่ละแบบดังกล่าวมีผลต่อการเคลื่อนที่ของโมเลกุลโปรตีนผ่านรู (pore หรือ seive) ของเจล (gel) โดยมีแรงต้านทานที่เกิดเนื่องจากความหนืดของเจลเอง

จากที่กล่าวมาข้างต้นถ้าให้โปรตีนที่มีประจุสุทธิ Q และขนาดรัศมีของโมเลกุล r ผ่านเจลที่มีความหนืด n ภายใต้สนามไฟฟ้าขนาด d แล้ว ความเร็วของการเคลื่อนที่ u หาได้จากสมการ

$$u = \frac{Qd}{4\pi r^2 n}$$

ซึ่งจะเห็นได้ว่าความเร็วของการเคลื่อนที่แปรผันตรงกับขนาดประจุสุทธิของโปรตีนและขนาดของสนามไฟฟ้า แต่แปรผกผันต่อขนาดของโมเลกุลโปรตีนและความหนืดของเจล

องค์ประกอบที่สำคัญของ electrophoresis ไม่ว่าจะเป็นเทคนิคใดก็ตามมีอยู่ 3 ส่วนคือ สนามไฟฟ้า เจลและบัฟเฟอร์ที่มีประจุ ขั้วไฟฟ้าจะต่อเข้ากับเจลที่ด้านตรงข้ามเพื่อสร้างสนามไฟฟ้าโดยมีบัฟเฟอร์ที่มีประจุเป็นตัวนำไฟฟ้าระหว่างขั้วไฟฟ้าทั้งสอง โปรตีนที่มีประจุบวกจะเคลื่อนเข้าขั้วลบ แต่โปรตีนที่มีประจุลบจะเคลื่อนเข้าหาขั้วบวก เจลจะช่วยในการแยกโปรตีนที่มีโครงสร้างต่างกันออกจากกันได้ด้วย ทำให้โปรตีนที่มีประจุและโครงสร้างต่างกันแยกออกจากกันได้ภายหลังเสร็จสิ้นการทำ eletrophoresis ในการตรวจดูตำแหน่งของโปรตีน (ไอโซไซม์) แต่ละแบบจะใช้เทคนิคทาง histochemistry โดยการนำเจลที่มีไอโซไซม์อยู่มาย้อมด้วยสีจำเพาะ จะปรากฏแถบที่ไอโซไซม์นั้น ๆ อยู่ให้เห็นได้ แถบตามแนวการเคลื่อนที่ของกระแสไฟฟ้าที่

มีตำแหน่งของไอโซไซม์ปรากฏให้เห็นเรียกว่า "ไซโมแกรม" (zymogram) (Hunter and Markert, 1957)

บัฟเฟอร์ที่มีประจุจะทำหน้าที่หลาย ๆ อย่าง ได้แก่ 1) ควบคุมไม่ให้ pH เปลี่ยนแปลง 2) เป็นตัวกลางให้กระแสไฟฟ้าไหลผ่านเจล 3) ลดอันตรกิริยา (interaction) ระหว่างประจุของเจลกับประจุของโปรตีนซึ่งอาจมีผลทำให้ประจุของโปรตีนเปลี่ยนแปลงไปได้ 4) ช่วยให้ไอโซไซม์มีความเสถียร และ 5) เป็นแหล่งของตัวเร่งปฏิริยา (catalyst) ที่จำเป็นของไอโซไซม์ เช่น Mg^{2+} เป็นต้น

หลักฐานที่ได้จาก isozyme electrophoresis มีทั้งข้อได้เปรียบและข้อจำกัดซึ่งพอกล่าวได้ดังนี้ (Gottlieb, 1977)

ข้อได้เปรียบ

1) เป็นการเปรียบเทียบในระดับจีนของสิ่งมีชีวิต ซึ่งขจัดปัญหาที่เกิดจากปรากฏการณ์ที่เรียกว่า convergence และหน้าที่การทำงานที่เหมือนกัน (functional correlation) ซึ่งมักประสบปัญหาดังกล่าวในการศึกษาลักษณะทางสัณฐานวิทยา

2) มีความถูกต้องและให้ค่าเกี่ยวกับจำนวนและชนิดของไอโซไซม์โดยตรงซึ่งสามารถใช้ข้อมูลดังกล่าวได้อย่างแท้จริง

3) การเปรียบเทียบจะกระทำกับเอนไซม์ที่มีอยู่ภายในสิ่งมีชีวิตและมีผลกระทบจากสภาพแวดล้อมน้อยมาก

4) ขจัดปัญหาเกี่ยวกับการให้ความสำคัญกับลักษณะที่ผู้ศึกษาคิดว่าสำคัญต่อการจัดกลุ่ม

ข้อจำกัด

1) เอนไซม์ที่ทำการศึกษายังมีจำนวนน้อยและไม่อาจเป็นตัวแทนของเอนไซม์โดยทั่วไปของสิ่งมีชีวิตได้เพราะเอนไซม์ที่ทำการศึกษาเป็นเพียงบางส่วนของขบวนการเมตาบอลิซึม

2) เนื่องจากกรดอะมิโน 1 ตัวอาจมีรหัสพันธุกรรมมากกว่า 1 รหัสได้ เมื่อมีการแทนที่เบสของรหัสพันธุกรรมเกิดขึ้น พบว่ามีเพียง 30 เปอร์เซ็นต์ของการแทนที่ของเบสที่ทำให้เกิดการแทนที่ของกรดอะมิโนซึ่งจะทำให้การเคลื่อนที่ของโมเลกุลไอโซไซม์ขณะทำ electrophoresis เปลี่ยนแปลงไป

3) ไอโซไซม์ที่สร้างจากจีนตำแหน่งเดียวกันแต่ต่างอัลลีลกันและมีการเคลื่อนที่ขณะทำ electrophoresis เหมือนกัน ไม่จำเป็นว่าจะมีลำดับของกรดอะมิโนเหมือนกัน นอกจากนี้ยังพบว่าไอโซไซม์มากกว่า 1 แบบมีการเคลื่อนที่ขณะทำ electrophoresis เหมือนกัน

4) หลักฐานจาก electrophoresis ไม่อาจบอกได้ว่ามีกรดอะมิโนต่างกันเป็นจำนวนเท่าใดหรือเกิด mutation กี่ขั้นตอน การเคลื่อนที่ของไอโซไซม์ขณะทำ electrophoresis ที่ต่างกันไม่สามารถให้รายละเอียดว่าเกิดการแทนที่ของเบสเพียง 1 เบสหรือมากกว่า 1

เบส ดังนั้นจึงไม่สามารถบอกได้ว่าสิ่งมีชีวิต 2 ชนิด ที่มีไอโซไซม์ที่สร้างจากจีนตำแหน่งเดียวกัน แต่ต่างอัลลีลกันนั้นมีความแตกต่างกันมากน้อยเพียงใด

นอกจากนี้ Triest (personal communication) ได้กล่าวเสริมถึงข้อได้เปรียบและข้อจำกัดของการศึกษา isozyme electrophoresis พอสรุปได้ดังนี้

ข้อได้เปรียบ

1) สามารถประยุกต์ใช้ในวงกว้างนับตั้งแต่ศึกษาในพืชแต่ละต้นไปจนถึงระดับพืชทั้งสกุล ประชากรพันธุศาสตร์ และการศึกษาอนุกรมวิธาน

2) ข้อมูลอัลโลไซม์ซึ่งอาจเป็นข้อมูลที่ต่อเนื่อง (ความถี่ของอัลลีล) หรือไม่ต่อเนื่อง (การมีหรือไม่มีของอัลลีลที่ทำการศึกษา) เป็นประโยชน์ต่อการศึกษาทางด้าน cladistics และ phenetics

ข้อจำกัด

1) อัลโลไซม์ที่ทำการศึกษามีความผันแปรไม่มากนัก

2) บางอัลลีลที่แตกต่างกันที่รหัสพันธุกรรมแต่ไม่แสดงออกในรูปของโปรตีน

3) มีผลกระทบเนื่องจากสภาพแวดล้อมหรือสรีรวิทยาภายในต้นพืช

หลักฐานที่ได้จาก isozyme electrophoresis สามารถนำมาใช้ในการหากำเนิดของพืชได้ ดังเช่น Gottlieb (1974) ศึกษาไอโซไซม์ 8 ระบบใน Clarkia lingulata โดยเปรียบเทียบกับ C. biloba (ssp. biloba และ ssp. australis) และ C. dudleyana เมื่อหาความเหมือนทางพันธุกรรม (genetic identity, I) ของพืชทั้ง 3 ชนิดจากไซโมแกรมทั้ง 8 ระบบพบว่า ค่า I ระหว่าง C. lingulata และ C. biloba สูงกว่าค่า I ระหว่าง C. lingulata กับ C. dudleyana ประกอบกับลักษณะภายนอกของ C. lingulata และ C. biloba มีความคล้ายคลึงกันมาก จึงสรุปว่า C. lingulata มีบรรพบุรุษเป็น C. biloba

Shen (1972) ใช้ isozyme electrophoresis เพื่อหากำเนิดของ Nicotina tabacum Linn. โดยศึกษาเอนไซม์ 8 ระบบใน N. tobacum (สายพันธุ์ Burley 21 กับ สายพันธุ์ Hicks), N. sylvestris, N. otophora, N. tomentosiformis, N. sylvestris x N. tomemtosiformis, N. sylvestris x N. otophora, N. sylvestris x Burley 21, และ amphidiploid ของ Kosnoff ที่เกิดจาก N. sylvestris x N. tomemtosiformis พร้อมทั้งหาค่าความเหมือนของพืชที่นำมาศึกษาจากไซโมแกรมที่ได้ ผลการศึกษาพบว่า N. sylvestris x N. tomemtosiformis, N.

tobacum และ amphidiploid ของ Kosnoff มีค่าความเหมือนสูงมาก ซึ่งแสดงให้เห็นว่า N. sylvestris และ N. tomemtosiformis เป็นบรรพบุรุษของ N. tobacum

นอกจากนี้ยังมีการประยุกต์ใช้ isozyme electrophoresis ในการตรวจหาสายพันธุ์พืชหรือการตรวจสถานะของหน่วยทางอนุกรมวิธานของพืช ดังเช่น Oliver and Ryjon (1980) ศึกษาเปรียบเทียบไอโซไซม์ระบบ esterase จากใบ ดอก และเมล็ดที่ยังไม่แก่ของ Muscari atlanticum (Linn.) Miller ที่เป็น ต้นเตตราพลอยด์ (4x) ต้นเพนตราพลอยด์ (5x) และต้นเฮกซาพลอยด์ (6x) สรุปได้ว่าสามารถจำแนกต้น 4x ออกจากต้น 5x และ 6x โดยใช้ไซโมแกรมที่ได้จากส่วนของดอกและเมล็ดที่ยังไม่แก่ แต่ไม่อาจแยกต้น 5x ออกจากต้น 6x ไม่ว่าจะพิจารณาไซโมแกรมของใบ ดอก หรือเมล็ดที่ยังไม่แก่

Torres, Soost, and Diedenhofen (1978) ศึกษาไอโซไซม์ 3 ระบบในพืชสกุล Citrus และพืชที่มีความสัมพันธ์กัน ได้ผลว่า สามารถใช้ข้อมูลจากไซโมแกรมของไอโซไซม์ทั้งสามระบบมาใช้ในการจำแนกต้นพืชที่เกิดการผสมระหว่างต้นพ่อต้นแม่ออกจากต้นพืชที่เกิดจากเนื้อเยื่อนิวเซลลัส (nucellus) ซึ่งเป็นประโยชน์ต่อการปรับปรุงพันธุ์ส้มมาก

Quiros (1980) ได้ใช้เทคนิค electrophoresis ในการศึกษาไอโซไซม์ระบบ esterase, peroxidase, acid phosphatase และ leucine aminopeptidase ใน Medicago sativa Linn. เพื่อจำแนกต้นพันธุ์ที่นำมาแทนต้นพันธุ์เดิมที่ตายไปว่ามีพันธุกรรมเหมือนกันหรือไม่ พบว่า สามารถใช้ไซโมแกรมของ esterase, peroxidase และ acid phosphatase ในการตรวจสอบว่าต้นพันธุ์ใหม่เหมือนต้นพันธุ์เดิมหรือไม่ ทั้งนี้ถ้าไซโมแกรมของไอโซไซม์ระบบหนึ่งเหมือนกันแล้ว ไซโมแกรมของไอโซไซม์อีกระบบจะต่างกัน

พืชสกุล Typha 4 ชนิด คือ T. latifolia Linn., T. angustifolia Linn., T. domigensis Pers. และ T. glauca Godr. มีความผันแปรค่อนข้างสูง ทำให้การตรวจหาชื่อผิดได้ นอกจากนี้ยังเชื่ออีกว่า T. glauca เป็นลูกผสมของ T. latifolia กับ T. angustifolia หรือ T. latifolia กับ T. domingensis Sharitz et al. (1980) ได้ศึกษาไซโมแกรมของไอโซไซม์ 10 ระบบเพื่อใช้ในการจำกัดขอบเขตของแต่ละชนิดได้ผลว่า ไซโมแกรมของไอโซไซม์ของ T. angustifolia, T. domigensis และ T. latifolia มีความแตกต่างกัน และสามารถใช้ในการตรวจหาชื่อได้ในกรณีที่ลักษณะสัณฐานวิทยาใช้ในการหาชื่อไม่ได้ ส่วนไซโมแกรมของไอโซไซม์ของ T. glauca ไม่สามารถสรุปได้ว่าเป็นลูกผสมของคู่ผสมใด แต่พบว่ามีความคล้ายคลึงกับไซโมแกรมของไอโอไซม์ของ T. latifolia และ T. angustifolia อย่างไรก็ตามยังมีความแตกต่างระหว่างไซโมแกรมของไอโอไซม์

ของ T. glauca กับอีก 3 ชนิด และเพียงพอที่จะกล่าวได้ว่า T. glauca เป็นพืชอีกชนิดที่มีอยู่จริง การตรวจสอบความสัมพันธ์ของพืชทั้ง 4 ชนิด พบว่า T. glauca มีความใกล้ชิดกับ T. latifolia มากกว่าอีก 2 ชนิด

Green et al. (1981) ศึกษาไอโซไซม์ 4 ระบบใน Stenotaphrum secundatum (Walt.) Kuntze จำนวน 28 clone ที่ใช้ปลูกในสนามหญ้า พบว่า ไอโซไซม์ระบบ ADP glucose pyrophosphorylase และ/หรือ UDP glucose pyrophosphorylase สามารถใช้แบ่ง S. secundatum เป็น 2 กลุ่มคือ กลุ่ม blue green S. secundatum และ yellow green S. secundatum และสามารถใช้ไอโซไซม์ระบบ Alcohol Dehydrogenase และ Acid Phophatase ในการจำแนก clone ที่อยู่ในกลุ่มย่อยทั้งสองได้

Cardy and Kannenberg (1982) ได้ทำการศึกษาอัลโลไซม์ของไอโซไซม์ 12 ระบบในข้าวโพดที่ใช้เพาะปลูก 113 line สรุปได้ว่า สามารถใช้ความผันแปรของอัลโลไซม์ของข้าวโพดในการตรวจสอบ line ของข้าวโพดได้

สมิต บุญเสริมสุขและประวิทย์ จิตตจำนงค์ (2532) ศึกษาไอโซไซม์ระบบ peroxidase ของพืชสกุล Azadirachta spp. 3 ชนิด ในประเทศไทย คือ สะเดาอินเดีย (A. indica Juss.), สะเดาไทย (A. indica var. siamensis Valeton) และไม้เทียม (A. excelsa Jack) ทั้งนี้ยืนยันการแยกไม้เทียมเป็น taxon ที่มีอยู่จริง ผลการศึกษาพบว่าไซโมแกรมของพืชทั้ง 3 ชนิด มีความแตกต่างกันอย่างเห็นได้ชัด แม้ว่าจะมีความผันแปรของลักษณะไซโมแกรมของไอโซไซม์ที่ปรากฏบ้างก็ตาม จึงสรุปได้ว่าไม้เทียมเป็นพืชอีกชนิดที่มีอยู่จริง

จากที่กล่าวมา (numerical taxonomy และ isozyme electrophoresis) ได้มีการนำวิธีการทั้งสองมาใช้ในการศึกษาทางอนุกรมวิธานดังจะขอกล่าวดังนี้

Wolff and Jefferies (1986a) ได้ศึกษาตรวจสอบสถานะของประชากรพืช Salicornia europaea Linn. (s.l.) ในภาคตะวันออกของทวีปอเมริกาเหนือ ซึ่งมี 7 ชนิด โดย 4 ชนิดเป็นดิพลอยด์ (diploid) ส่วนอีก 3 ชนิดเป็นเตตราพลอยด์ (tetraploid) การตรวจสอบชื่อของแต่ละชนิดค่อนข้างยุ่งยากเพราะดอกมีการลดรูปลงไปมาก นอกจากนี้ยังมีการแยกเอาพืชที่มีการกระจายอยู่บริเวณตอนกลางของการกระจายพันธุ์เป็นอีกชนิดหนึ่ง คือ S. rubra Nels. ดังนั้นในการศึกษาจึงมุ่งตรวจสอบสถานะของทั้ง 7 ชนิดภายใน S. europaea (s.l.) และการมีอยู่จริงของ S. rubra การวิเคราะห์ลักษณะสัณฐานวิทยา

ด้วยเทคนิคการวิเคราะห์จัดจำแนกโดยกำหนดกลุ่มจากการกระจายพันธุ์และระดับพลอยดี พบว่า S. europaea (s.l.) ประกอบด้วย 3 กลุ่ม คือ (I) เตตราพลอยด์ที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติก (II) ดิพพลอยด์ที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติกและอ่าวเจมส์ และ (III) ดิพพลอยด์ที่อยู่ตามอ่าวฮัดสัน โดยลักษณะที่แยกพืชทั้ง 3 กลุ่มออกจากกัน ได้แก่ ความกว้างของเยื่อที่พันรอบข้อของปล้องที่มีดอก ความยาวของอับเรณูและความกว้างของปล้องที่มีดอก การวิเคราะห์จัดจำแนก S. europaea (s.l.) ที่เป็นดิพพลอยด์และ S. rubra พบว่า ประกอบด้วย 3 กลุ่ม คือ (A) พืชที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติกและอ่าวเจมส์ (B) พืชที่อยู่ตามอ่าวฮัดสัน และ (C) S. rubra ทั้งนี้การทำนายกลุ่มจากสมการของพืชทั้ง 3 กลุ่มถูกต้องมากกว่า 92 เปอร์เซ็นต์ โดยลักษณะที่แยกพืชทั้ง 3 กลุ่มออกจากกัน ได้แก่ ความยาวของอับเรณู และความกว้างตอนกลางและตอนบนของปล้องที่ดอก การศึกษาไซโมแกรมของไอโซไซม์ 21 ระบบ พบว่า มีเพียง 5 ระบบที่มีไซโมแกรมแตกต่างกัน และสามารถแยกพืชออกเป็น 3 กลุ่ม คือ เตตราพลอยด์ที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติก, ดิพพลอยด์ S. europaea (s.l.) และ S. rubra ในพืชที่เป็นดิพพลอยด์ พบว่า พืชที่อยู่ตามอ่าวฮัดสันและอ่าวเจมส์มีความถี่ยีนที่ตำแหน่งเดียวกันเท่ากัน และมีความแตกต่างไปจากพืชที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติก เมื่อกำหนดกลุ่มโดยใช้ข้อมูลจากไซโมแกรมของไอโซไซม์ แล้ววิเคราะห์ด้วยการวิเคราะห์จัดจำแนก พบว่า ใน S. europaea (s.l.) ประกอบด้วย 3 กลุ่มซึ่งเหมือนกับการวิเคราะห์ข้างต้น แต่ลักษณะที่แยกพืชทั้ง 3 กลุ่มออกจากกันได้แก่ความกว้างของเยื่อที่พันรอบข้อของปล้องที่มีดอก ความยาวของอับเรณูและความกว้างของฐานของปล้องที่มีดอก เป็นที่สังเกตว่าพืชที่อยู่ตามอ่าวเจมส์ถูกจัดไว้ในกลุ่มเดียวกับพืชที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติกแม้ว่าไซโมแกรมของไอโซไซม์จะมีความถี่ยีนต่างกัน การวิเคราะห์จัดจำแนกในพืชที่เป็นดิพพลอยด์ที่กำหนดกลุ่มโดยไซโมแกรมของไอโซไซม์ พบว่า ประกอบด้วย 3 กลุ่มเช่นเดียวกับข้างต้น และพืชที่อยู่ตามอ่าวเจมส์ถูกจัดไว้ในกลุ่มเดียวกับพืชที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติกแม้ว่าไซโมแกรมของไอโซไซม์จะมีความถี่ยีนต่างกัน ลักษณะที่แยกพืชทั้งสามออกจากกัน ได้แก่ ความยาวของอับเรณู ความกว้างตอนกลางและตอนบนของปล้องที่มีดอก จากที่กล่าวมาแสดงว่า S. europaea Linn. (s.l.) ประกอบด้วย 3 ชนิดคือ (A) เตตราพลอยด์ที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติก (B) ดิพพลอยด์ที่อยู่ตามชายฝั่งมหาสมุทรแอตแลนติกและอ่าวเจมส์ และ (C) ดิพพลอยด์ที่อยู่ตามอ่าวฮัดสัน และ S. rubra เป็นพืชอีกชนิดที่มีอยู่จริงโดยมีการกระจายบริเวณตอนกลางของการกระจายพันธุ์ของ S. europaea

Wolff and Jefferies (1986b) ได้ศึกษาเปรียบเทียบพืช Salicornia rubra และ S. europaea Linn. (s.l.) ที่เป็นดิพพลอยด์ในภาคตะวันตกเฉียงเหนือของทวีปอเมริกาเหนือ กับ S. europaea Linn. (s.s.) และ S. ramosissima จากทวีปยุโรป การแบ่งกลุ่มประชากรของ S. europaea Linn. (s.l.) ออกเป็น 3 กลุ่ม คือ (i) กลุ่มที่มีการ-

กระจายพันธุ์อยู่ตามชายฝั่งมหาสมุทรแอตแลนติก (ii) อ่าวเจมส์ และ (iii) อ่าวฮัดสัน การศึกษาด้วยวิธีการทาง numerical taxonomy โดยใช้เทคนิคการวิเคราะห์จัดจำแนก S. europaea (s.s.), S. ramosissima, S. europaea (s.l.) ทั้ง 3 กลุ่ม และ S. rubra การศึกษาตอนแรกศึกษาว่า S. europaea (s.s.) และ S. ramosissma เป็นแต่ละชนิด (species) หรือไม่ ได้ผลว่า S. europaea (s.s.) และ S. ramosissima แยกเป็นแต่ละชนิดโดยลักษณะที่แยกพืชทั้งสองออกจากกันคือความยาวของอับเรณู ความกว้างส่วนยอดของปล้องที่มีดอกและความยาวของ lateral perianth การศึกษาในตอนที่สองใช้พืชทั้ง 4 ชนิดได้ผลว่าการกระจายของคะแนนจัดจำแนกประกอบด้วย 5 กลุ่ม คือ (1) S. europaea (s.s.), (2) S. ramosissima, (3) S. rubra, (4) S. europaea (s.l.) (i) กับ (ii) และ (5) S. europaea (s.l.) (iii) โดยมีลักษณะความยาวอับเรณู ความกว้างตอนกลางของปล้องที่มีดอก ความกว้างของเยื้อที่ผันข้อของปล้องที่มีดอก และความยาวของ lateral perianth เป็นลักษณะที่แยกพืชทั้ง 5 กลุ่มออกจากกัน ทั้งนี้การศึกษาโดยวิธี numerical taxonomy ทั้งสองตอนมีเปอร์เซ็นต์การทำนายกลุ่มจากสมการถูกต้องสูงถึงร้อยละ 99 การศึกษาไซโมแกรมของไอโซไซม์ 21 ระบบในพืชทั้ง 4 ชนิด พบว่า ไซโมแกรมของไอโซไซม์เพียง 5 ระบบเท่านั้นที่มีความแตกต่างกัน เมื่อหาความถี่ของแต่ละอัลลีลพบว่า S. europaea (s.s.) และ S. ramosissima มีความถี่ยีนเพียง 2 อัลลีลที่เท่ากัน ส่วนอัลลีลอื่นมีความถี่ต่างกัน และทั้ง 2 ชนิดมีความถี่อัลลีลที่ตำแหน่งต่าง ๆ แตกต่างไปจาก S. rubra และ S. europaea (s.l.) ด้วย โดยเฉพาะ 2 ชนิดหลังมีความถี่ของอัลลีลต่าง ๆ ไม่เท่ากัน ใน S. europaea (s.l.) พบว่า ประชากร (i) มีความถี่อัลลีลต่าง ๆ แตกต่างไปจากประชากร (ii) และ (iii) ซึ่งมีความถี่ของอัลลีลต่าง ๆ เท่ากัน จากข้อมูลความถี่ของอัลลีลสามารถแบ่งประชากรของพืชทั้ง 4 ชนิดได้เป็น 5 กลุ่มคือ S. europaea (s.s.), S. ramosissima, S. rubra, S. europaea (s.l.) กลุ่ม (i) และ S. europaea (s.l.) กลุ่ม (ii) และ (iii) แล้วศึกษาโดยวิธีการทาง numerical taxonomy พบว่า S. europaea (s.s.), S. ramosissima, S. rubra และ S. europaea (s.l.) แยกกันเป็นแต่ละชนิด แต่ใน S. europaea (s.l.) นั้นประกอบด้วย 2 ประชากรคือ (1) กลุ่ม (i) และกลุ่ม (ii) และ (2) กลุ่ม (iii) เหมือนการวิเคราะห์ในตอนที่สอง นั้นคือจากผลการศึกษาด้วยวิธีการของ numerical taxonomy และไอโซไซม์ของพืชทั้ง 4 ชนิด สรุปได้ว่า S. europaea (s.s.), S. ramosissima และ S. rubra แยกเป็นแต่ละชนิดแล้ว และควรแยกประชากร (i) (ii) และ (iii) ใน S. europaea (s.l.) เป็นอีกชนิดหนึ่ง โดยกลุ่ม (i) กับกลุ่ม (ii) เป็น S. borealis Wolff & Jefferies และกลุ่ม (iii) เป็น S. maritima Wolff & Jefferies ส่วน S. europaea Linn. (s.l.) ที่กระจายพันธุ์ในพื้นที่อื่นยังคงอยู่ในชนิดเดิม

Microscopic study

การศึกษาอนุกรมวิธานสามารถนำเอาลักษณะสัณฐานวิทยาของเมล็ดและลวดลายบนผิวของเมล็ดมาใช้ได้เพราะเป็นลักษณะที่ค่อนข้างคงที่ในพืชแต่ละชนิดและมีความสำคัญต่อการหาความสัมพันธ์ทางสายพันธุ์ของพืชที่ทำการศึกษาได้ ดังตัวอย่างที่จะกล่าวถึงดังนี้

Chuang and Heckard (1972) ได้ทำการศึกษาลวดลายของเปลือกหุ้มเมล็ดของพืชสกุล Cordylanthus (วงศ์ Scrophulariaceae) จำนวน 23 ชนิด (species) ด้วยกล้องจุลทรรศน์อิเลกตรอนแบบส่องกราด เมล็ดของพืชที่นำมาศึกษาอยู่ใน 3 section คือ section Dicranostegia 1 ชนิด, section Hemistegia 6 ชนิด และ section Cordylanthus 16 ชนิด ผลการศึกษาพบว่า ลักษณะลวดลายบนผิวเมล็ดมีความแตกต่างกัน จำแนกได้ 4 แบบ คือ แบบที่ 1 เป็นสันที่เป็นลอนสูง ๆ ต่ำ ๆ แบบที่ 2 เป็นร่างแหที่ช่องลึก แบบที่ 3 เป็นร่างแหที่ช่องตื้น และแบบที่ 4 การเรียงตัวของสันคดเคี้ยวคล้ายคลื่น สำหรับแบบที่ 2 และแบบที่ 3 นั้น Chuang and Heckard ให้ความเห็นว่าคือแบบเดียวกัน เพียงแต่ต่างกันที่ช่องลึกหรือตื้นเท่านั้น เมล็ดของพืชใน section Dicronostegia มีลวดลายแบบที่ 1 เมล็ดของพืชใน section Hemistegia มีลวดลายแบบที่ 2 ยกเว้น C. palmatus (Ferris) Macbr. ที่มีลวดลายเป็นแบบที่ 1 แต่เดิมนั้น C. palmatus ถูกจัดว่าเป็นชนิดเดียวกับ C. hispidus Pennell แต่จากลวดลายผิวของเมล็ดของพืชทั้งสองแสดงว่า C. palmatus เป็นอีกชนิดหนึ่ง ส่วนเมล็ดของพืชใน section Cordylanthus มีลวดลายเป็นแบบที่ 2 3 และ 4 จากลวดลายบนผิวเมล็ดของพืชใน section นี้แสดงว่าน่าจะมีสายวิวัฒนาการของพืชใน section อย่างน้อย 2 สาย คือ สายแรกเป็นพืชที่มีลวดลายบนผิวเมล็ดเป็นแบบที่ 2 และ 3 ส่วนสายที่ 2 เป็นพืชที่มีลวดลายบนผิวเมล็ดเป็นแบบที่ 4

Whiffin and Tomb (1972) ศึกษาลักษณะผิวของเมล็ดพืชวงศ์ Melastomataceae 5 tribe คือ Microlicieae, Tibouchineae, Bertolonieae , Merianieae และ Rhixieae พบว่า รูปร่างของเมล็ดแบ่งได้ 5 แบบคือ Microlicioid (Mi), Tibouchinioid (Ti), Rhixioid (R), Merianioid (Me) และ Bertolonioid (B) พืชที่อยู่ใน tribe Bertolonieae มีรูปร่างเมล็ดเป็นแบบ B และรูปร่างของเมล็ดของพืชใน tribe Merianieae เป็นแบบ Me ใน tribe Rhixieae เมล็ดมีรูปร่างเป็นแบบ R ซึ่งจะพบเมล็ดรูปร่างเดียวกันนี้ในเมล็ดของพืชสกุล Arthrostma ซึ่งอยู่ใน tribe Tibouchineae แสดงว่า พืชสกุล Arthrostema มีความใกล้ชิดกับพืชใน tribe Rhixieae มากกว่าพืชที่อยู่ใน tribe เดียวกัน ส่วนพืชใน tribe Tribouchineae อีก 7 สกุลมีเมล็ดรูปร่างแบบ T

และมีสัณฐานวิทยาคล้ายคลึงกันมาก แสดงว่าพืชทั้ง 7 สกุลมีความใกล้ชิดกันมาก อย่างไรก็ตาม ถึงแม้ว่าเมล็ดของพืชสกุล Monochaetum ซึ่งอยู่ใน tribe Tibouchneae จะมีรูปร่างเป็นแบบ T แต่มีความแตกต่างไปจากลักษณะเมล็ดของพืชใน 7 สกุลที่อยู่ใน tribe เดียวกัน นอกจากนี้ เมล็ดของพืชในสกุล Chaetolepis ซึ่งอยู่ใน tribe Rhixieae แต่มีรูปร่างเมล็ดเป็นแบบ T และมีลักษณะคล้ายคลึงกับเมล็ดของพืชในสกุล Monochaetum แสดงว่าพืชทั้งสองสกุลนี้มีความใกล้ชิดกันมากและควรมีการจัดจำแนกพืชทั้งสองสกุลใหม่ และรูปร่างเมล็ดของพืชใน tribe Microlicieae เป็นแบบ Mi ซึ่งจะพบรูปร่างเมล็ดแบบนี้ในเมล็ดของพืชใน tribe Tibouchineae ด้วย โดยสรุปต้องมีการศึกษาพืชทั้ง 5 tribe นี้ใหม่เพื่อจัดกลุ่มให้ถูกต้องยิ่งขึ้น

Menapace (1991) ศึกษาเมล็ดของพืชสกุล Eleocharis ซึ่งเป็นพืชที่มีลักษณะสัณฐานวิทยาของต้นและดอกลดรูปลงไปมาก ทำให้แยกพืชแต่ละชนิดได้ลำบาก Menapace ใช้ลักษณะเมล็ดของพืชสกุลนี้เพื่อตรวจสอบว่าจะสามารถใช้เป็นหลักฐานในการตรวจหาชนิดได้หรือไม่ ผลการศึกษาพบว่า จากหลักฐานลักษณะรูปร่างและผิวของเมล็ดของพืชในสกุลนี้สามารถใช้แยกพืชแต่ละ series ออกจากกันได้ ในพืชจำนวน 9 series มีเพียง 6 series เท่านั้นที่เป็น series ที่มีอยู่จริงในธรรมชาติ อีก 3 series คือ Pauciforae, Texuissimae และ Maculosae ยังไม่แน่ใจว่าเป็น series ที่มีอยู่จริง กล่าวคือ พืชใน series อื่นไม่มี silica body ที่เมล็ด แต่พืชใน series Pauciforae มี silica body ที่เมล็ดซึ่งเหมือนกับเมล็ดของพืชในสกุล Carex, Scirpus, Cyperus, Eriophorume ดังนั้นพืชใน series Pauciforae น่าจะเป็นพืชในสกุลใดสกุลหนึ่งใน 4 สกุลที่กล่าวมา ส่วนพืชใน serie Maculosae ซึ่งประกอบด้วย 2 subseries แต่พืชทั้งสอง subseries มีลักษณะเมล็ดที่เหมือนกันมาก จึงไม่น่าเชื่อว่าควรมี 2 subseries ใน series นี้ สำหรับ series Tenuissimae ซึ่งเกิดจากการรวม section Chaetarieae และ Leiacarpaea พบว่าพืชในแต่ละ section มีลักษณะของเมล็ดแตกต่างกัน จึงควรแยกพืชใน series นี้ออกเป็นแต่ละ section

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย