1.1.2. Crop origin, evolution, diversity, collection, domestication

Crop domestication (Harlan’ paper)

– Man domesticated plants, or
– Plants domesticated man

– Transition from hunter-gatherer to agricultural man
  • Is it really the next step in the evolution?
  • Why some hunter-gatherers prevailed until modern times?
What did gatherers eat?
(valid data mostly limited to the new world and Africa)

• Grass seed – ca. 60 spp.
  – Potential cereals, e.g. Panicum favourite of gatherers all over the world especially Australia, NA; wild rice (Zizania aquatica) in NA
  – Some wild grass seed has similar or superior nutritional value than modern wheat.

Legumes – ca. 50 spp.

  – Whole pods, seeds or other tissue
  
  – Learned to detoxify poisonous tissues, or use poison to stunning fish, stupefying emus or poison arrow
  
  – Some genera widely used – Acacia used in Australia, Africa and Americas.
    • Tephrosia spp. used for fish poison on five continents.
Roots and tubers – ca. 90 spp.

- Widely used by all gatherers

- Genus *Discorea* has approx. 600 spp.
  - Many of them produce edible tubers (or after detoxification)

- Wild yam – India, SE Asia, Australia, tropical America.

- Onions, tuberous legumes, Solanaceae, etc...

Cyperus rotundus (nutgrass)

- used for food in N. America, Asia, Africa, Australia and Europe.
Oil seeds – ca. 60 spp.
- In spite of access to animal fat, gatherer-hunter used various sources of vegetable oil
  - Oil palms, coconut, avocado, pistachio, olives
  - Seeds of Compositae, Cruciferae, Cucurbitaceae partly for their oil
  - *Linum* and *Sesamum*

- Fruits and nuts >500 spp.
  - Too many to list – obvious various nuts used everywhere

- Vegetable and spices >600 spp.
  - Too many to list
    - Solanaceae found on every continent, various species used for food, drug (recreational use) or medicine
    - Cucurbitaceae important because of their abundance in some areas – wild watermelon in southern Africa (it may served only as a source of water).
Strong evidence that the gatherers understood the life cycle of plants

- Planting seed of non-domesticated plants (to thicken the stand of desirable plants)
- Various cultivation practices such as burning the vegetation and irrigation
- Botanical knowledge – detoxification of poisonous plants

Origins of Agriculture not clear
(why man started growing crops)
- Harlan’s paper gives an excellent overview but still unclear
  - Agriculture in mythologies of various cultures (always divine origin or status)
  - Domestication for religious reasons
  - Crowding – due to climate change and reduced gathering range
  - Agriculture as discovery
  - Agriculture as an extension of gathering
Geography of Plant Domestication
Alphonse de Candolle – Origin of Cultivated Plants (1886)
Nikolai Ivanovich Vavilov – On the Origin of Cultivated Plants (1926)

- Determine the origin by analyzing the pattern of variation
- The geographic region with the greatest genetic diversity = region of origin
- Eight (8) centers proposed, with some subcenters
- More philosophical doctrine and very little data to support the 8 centres
- Launched expeditions around the world to gather all the useful germplasm of all crops that had potential for the Soviet Union

Nikolai Ivanovich Vavilov
• Proposed that the centre of origin of species coincide with the areas of greatest diversity for the species.

• Later recognized secondary centres of origin (diversity)

• Law of homologous series in variation –
  − variation in one species is found in another closely related species. Used to predict characters remaining to be discovered or developed. I.e. if a trait is present in one species, it is likely that it could be either found in a related species, or developed through mutation, interspecific crosses etc...
  − Herbicide tolerance in weeds or some crops.
Vavilov's centers of origin of cultivated plants

1. China
2. India (2a – Indochina)
3. Central Asia
4. Near East
5. Mediterranean
6. Ethiopia
7. S. Mexico - Central America
8. South America a) Chile b) Brasil
1. **China** - The largest independent center - total of 136 endemic plants

- Soybean, several forms of millet, hull-less barley, Chinese yam, Chinese cabbage, onion, cucumber, pear, peach, apricot, cherry, walnut, sugarcane, opium poppy, hemp

2. **India** - 117 species

- Rice, chickpea, mung bean, cowpea, eggplant, cucumber, radish, yam, mango, orange, sugarcane, coconut palm, sesame, safflower, jute, kenaf, oriental cotton, hemp, black pepper, gum Arabic, indigo, cinnamon tree, bamboo

2a. **Sub-Center Indochina** - 55 species

- Banana, coconut palm, sugarcane, nutmeg, manila hemp
3 Central Asia (N. India, Afghanistan, Turkmenistan) - 43 species

- Wheat (T. Aestivum) and other hexaploid wheat, peas, lentils, chickpeas, mustard, flax, sesame, cotton, hemp, onion, garlic, spinach, carrot, pistachio, pear, almond, apple
  - Small number of species, but great variation

4 The Near East – 83 species

- Many Triticum (diplo- and tetraploid), two-row barley, rye, lentil, lupine, alfalfa, Persian clover, vetch, fig, pomegranate, apple, pear, cherry

5 Mediterranean Sea, coastal and adjacent regions – 84 endemic species

- Durum wheat, spelt (wheat), Mediterranean oats, sand oats, canarygrass, pea, lupine, white clover, crimson clover, flax, rapeseed, black mustard, olive, garden beet, cabbage, turnip, lettuce, asparagus, celery, chicory, parsnip, rhubarb, anise, thyme, peppermint, sage, hop.
6 Ethiopia – 38 species listed
• Barley, tetraploid wheat, grain sorghum, pearl millet, cowpea, flax, sesame, castor bean, coffee, indigo

7 Southern Mexico and Middle America
• Maize, cotton, cocoa, beans, pepper, squash

8 South America – 62 plant species, 2 sub-centers
• Potatoes (2n=24, 36, 60), starchy maize, lima bean, common bean, tomato, pumpkin, pepper, Egyptian cotton, cocoa, guava, tobacco

8a Sub-Center Chile
• Common potato (2n=48)

8b Sub-Center Brazil
• Peanuts, pineapple, Brazil nut, cashew, rubber tree
**Jack Harlan**

- Recognized that patterns of variability are more complex and more diffused than Vavilov originally imagined.
- Centers of diversity are not the same as the centers of origin.
- Many crops have centers of diversity, sometimes multiple.
- Some crops do not have centers of diversity.

**Evolutionary patterns (Harlan)**

- **Endemic** – crops that originated in a limited area and very little dispersal.
- **Semiendemic** – crops that originated in definable center and with limited dispersal.
- **Monocentric** – crops with definable center of origin and wide dispersal without secondary centers of diversity.
- **Oligocentric** – crops with definable center of origin, wide dispersal, and one or more secondary centers of diversity.
- **Noncentric** – crops whose pattern of variation suggest domestication over a wide area
  - Sorghum, rapeseed (?), common bean (*Phaseolus vulgaris* Linn.)
- **Microcenters**
Harlan’s centres of diversity

A1 Near East
- Wheat, Barley, Oats, Rye, Apple, Cherry, Date, Lettuce, Green Pea, Melon, Beet, Cabbage, Carrot, Onion

A2 Africa
- Sorghum, Millet, Cowpea, Yam, Cotton, Coffee

B1 China
- Rice, Buckwheat, Millet, Soybean, Cucumber (India), Peach, Tea

B2 Southeast Asia
- Rice, Banana, Orange, Grapefruit, lemon, Black Pepper, Cloves, Nutmeg, Coconut, Sugarcane

C1 Mesoamerica
- Maize, Bell Pepper, Common Bean, Squash, Tomato, Lima bean, Sweet potato, Cotton, Pineapple, Cacao, Sunflower

C2 South America
- Common bean, Peanut, Potato, Squash, Tobacco, Avocado, Cotton
Summary of Consensus

<table>
<thead>
<tr>
<th>Region</th>
<th>Crops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near East (Fertile Crescent)</td>
<td>Wheat and barley, flax, lentils, chickpea, figs, dates, grapes, olives, lettuce, onions, cabbage, carrots, cucumbers, and melons; fruits and nuts</td>
</tr>
<tr>
<td>Africa</td>
<td>Pearl millet, Guinea millet, African rice, sorghum, cowpea, Bambara groundnut, yam, oil palm, watermelon, okra</td>
</tr>
<tr>
<td>China</td>
<td>Japanese millet, rice, buckwheat, and soybean</td>
</tr>
<tr>
<td>South East Asia</td>
<td>Wet- and dryland rice, pigeon pea, mung bean, citrus fruits, coconut, taro, yams, banana, breadfruit, coconut, sugarcane</td>
</tr>
<tr>
<td>Mesoamerica &amp; North America</td>
<td>Maize, squash, common bean, lima bean, peppers, amaranth, sweet potato, sunflower</td>
</tr>
<tr>
<td>South America</td>
<td>Lowlands: cassava mid altitudes and uplands (Peru): potato, peanut, cotton, maize</td>
</tr>
</tbody>
</table>
An Ecological Approach to Geography of Domestication

Major climate or vegetation formation

1. Tundra and taiga – nothing
2. Temperate forests – fruits and nuts
3. Temperate prairies – sunflower
4. Temperate steppes – marginal but *Aegilops squarrosa* (contributed D genome to hexaploid wheat)
5. Mediterranean woodlands – many
6. Tropical forest – banana, orange, mango (on the forest margins)
7. Tropical savannas – many
8. Deserts – very few
9. Tropical highlands – some major corps (potato) and many minor ones (*arabica* coffee)
10. Sea coast – some important crops – coconut, cabbage, beet, radish

- 5 and 7 both have long dry seasons – generate annuals and plants that behave as annuals (why is it could be important?)

### Effects of photosynthetic pathway on plant adaptation to different environments, growth and productivity

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>C3 Species</th>
<th>C4 Species</th>
<th>CAM Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf anatomy</td>
<td>No distinct bundle sheath cells</td>
<td>Bundle sheath cells</td>
<td>No distinct bundle sheath cells</td>
</tr>
<tr>
<td>Stomata</td>
<td>Open during the day</td>
<td>Open during the day</td>
<td>Usually open at night and closed during the day</td>
</tr>
<tr>
<td>Transpiration ratio*</td>
<td>350 - 1000</td>
<td>150 - 300</td>
<td>50 to 100</td>
</tr>
<tr>
<td>First product</td>
<td>3-phosphoglyceric acid</td>
<td>oxaloacetic acid (converted to malic or aspartic acid)</td>
<td>oxaloacetic acid (converted to malic acid)</td>
</tr>
</tbody>
</table>

*The ratio kg water transpired per kg dry weight produced (low values indicate high water use efficiency)*
### Effects of photosynthetic pathway on plant adaptation to different environments, growth and productivity

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>C3 Species</th>
<th>C4 Species</th>
<th>CAM Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of photosynthesis</td>
<td>Entire leaf</td>
<td>Bundle sheath cells</td>
<td>Entire leaf</td>
</tr>
<tr>
<td><strong>Light response</strong></td>
<td>Saturated at half of full sunlight</td>
<td>Not saturated at full sunlight</td>
<td>Saturated at one fourth of full sunlight</td>
</tr>
<tr>
<td>Photorespiration &amp; CO₂ compensation point</td>
<td>Yes, 50 ppm</td>
<td>No, 10 ppm or less</td>
<td>Yes, 50 ppm in light</td>
</tr>
<tr>
<td>Photosynthesis rate, umoles m⁻² s⁻¹</td>
<td>6 to 40</td>
<td>14 - 64</td>
<td>1.5 to 6</td>
</tr>
<tr>
<td>Maximum growth rate, g dm⁻¹</td>
<td>34 - 39</td>
<td>50 - 54</td>
<td>~15, up to 20</td>
</tr>
<tr>
<td>Average productivity, ton ha⁻¹ yr⁻¹</td>
<td>~40</td>
<td>60 to 80</td>
<td>&gt;40</td>
</tr>
</tbody>
</table>

**Light saturation of a single leaf is indicated by failure of the CO₂ assimilation rate to increase with an increase in light intensity**

### Effects of photosynthetic pathway on plant adaptation to different environments, growth and productivity

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>C3 Species</th>
<th>C4 Species</th>
<th>CAM Species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Climatic adaptation</strong></td>
<td>Temperate to tropical</td>
<td>Tropical</td>
<td>Arid tropical to Mediterranean</td>
</tr>
<tr>
<td>Crops</td>
<td>rice, wheat, barley, soybean, peanut, potato, sweet potato, taro, banana, bean, most vegetables, beet, cabbage, sunflower, all fruit trees studied, etc</td>
<td>corn, sugarcane, sorghum, millets, tropical grasses, Chinese spinach (an amaranth)</td>
<td>pineapple, prickly pear cactus, many orchids, sissal and other agaves, other cactus, etc.</td>
</tr>
</tbody>
</table>
Variation Patterns and Implications for Plant Breeding

Typical Variation Patterns:

• **Wild population** are often highly variable, especially over a large area and/or ecological amplitude.

• **Landrace populations** are often balanced, integrated mixtures of genotypes adapted to the region and cultivation practices.

• **Weed population**, often derived from genetic interaction between wild and cultivated races.

Variation Patterns and Implications for Plant Breeding

Typical Variation Patterns:

• Microcenters in which enormous diversity is found in a restricted geographic area, usually due to genetic interaction between cultivated and/or spontaneous races.

• Secondary centers in which great variation has accumulated in certain special geographic regions, usually with considerable isolation from other regions for a long period of time.
Landrace populations

- Highly variable in appearance, but identifiable and usually have local names
- May have adaptation to a particular soil classification (heavy-light, wet-dry, warm-cold)
- May have utility differentiation (flour, malt, etc.)

Landrace populations

- Genetic variation considerable but not random
- Genotypes adapted to the local environment and to each other, i.e. an integrated unit and the components have adjusted to one another over generations.
- The composition is frequently deliberately manipulated by cultivation.
  - Non-random harvest of seed for planting the next generation
Class discussion:
How important is the knowledge of centres of origin / diversity to modern breeding and your crop in particular?
Plant Breeding Objectives

Plant breeding is often defined as **science**, **art** and **business** of improving plants for the human benefit (Bernardo, 2002)

**Science** – grounded on a theoretical and empirical body of knowledge, most notable genetics
  - Objective basis for deciding what parents to cross, selection method, progenies to keep
  - It does not follow the basic scientific method: hypothesis-test-theory-law

**Art** – subjective judgement in the design and implementation of a breeding program
  - Breeder’s eye – the intuition that one material is better than other

**Business** – requires investment of people, money and time
  - Return on investment regardless how the cost/benefit is measured
Goals of Plant Breeding

Two fundamental decisions are:

1. **What combination of traits to breeder for**
2. **What group of environments to breed for**
   (usually associated with markets)

**Increased yield – always a good idea**

the ultimate objective of most breeders, 
*commoditized* crops (*the more of the same*)

v.s. non-commoditized (*specialty*) crops – (*the more of the different*)

– Specific improvement such as disease resistance, lodging resistance etc...

– Generally greater physiological efficiency

– Difficult to differentiate between the two, especially retrospectively
Increase in harvestable component –
Most notable example - Sugar beets

- sugar content increased from <7% to >18%

- 175 years of breeding (~1%/year of breeding)

Protecting the yield

- Breeding for disease and insect resistance
  - The most dramatic and best recognized plant breeding effort
    - Wheat stem rust
    - *Helminthosporium maydis* in US corn
    - Canola resistance to blackleg (*Phoma lingam*)
      - crop wiped out in Canada and Australia
    - Vegetables in US and Europe
      - Lettuce
        - Downy mildew (*Bremia lactucae*); Currant Lettuce Aphid (*Nasonovia*)
      - Tomato
        - TMV, Cladosporium, Verticilium, Fusarium, Root-knot nematodes
• **Introduction of new crops and development of adapted varieties**
  - have made big economic impact in many areas/crops
    - Grain Sorghum
      - A tropical grass introduced to US 150 years ago
      - Initially confined to warmer parts of SW US
      - Earlier and earlier varieties have expanded the areas of cultivation (up to S. Dakota)
  - Canola in Western Canada
    - 15-20 years ago - mostly Swedish and Danish *B. napus* varieties required 110-120 days to mature
      - Market split 50% *B. napus*; 50% *B. rapa* (7-10 days earlier)
    - Now ~ 100% *B. napus* – varieties mature in 95-110 days

• **Breeding for abiotic stress tolerance**
  - Drought resistance
    - water use efficiency
  - Excess moisture (flood) tolerance
  - Heat tolerance
  - Cold tolerance
• **Improvement of agronomic characteristics**
  
  • Reducing the height

  – Allows higher inputs (N)

  – Allows mechanized harvest

• **Changing the utilization of the crop**

  – Energy crops (corn, soybean, canola, alfalfa) – focus on energy rather than e.g. protein content

  – Low linolenic acid soybean – high stability oil for use in industrial frying

  – Fresh-cut fruit market – the customer does not see the fruit any longer, other characteristics become important

  – Canola oil is one of the healthiest vegetable oils, developed from rapeseed – oil was used as an industrial lubricant (4-5 allele difference)
• Changing growing condition

  – Moving vegetable production from field to protected (greenhouse) environment
  – New diseases, insects and abiotic stress conditions

• Field crops

  – Shortening rotation
  – Zero or minimum tilling
  – Climate changes

• End use market driven traits

  – Aesthetically pleasing shape or color

  • Ornamental plants

  • Fruits and vegetable
Utility driven traits

• Mechanized harvest

• Size and shape to fit the packaging
  — e.g. canning or jarring
  — Increasing the yield without increase the size of the fruit??

• Pickling cucumber must fit the jar
• Cherry tomato = a half bite size

• Competitiveness of the market

  — Differentiation:
    • real improvement
    • new trait

  — Marketing driven objectives – need for a new product every 2, 3 or 5 years
Common questions
• The role of the breeder?

• Will it change with the new development?

• What competencies will be required from the breeder?
  – Genome sequences, bioinformatics, biostatistics,

Factors elements that influence the success of a plant breeding program are:

1. Row material (germplasm)

2. Skills (technical knowledge, experience)

3. Resources ($$, and everything you could buy – equipment, land etc..)
Comparison to a manufacturing process
(Bernardo 2010)

Skills and experience

Raw materials (Germplasm) ➔ Manufacturing (Inbred, Hybrid Development) ➔ Testing and quality control (Field Trials) ➔ Product (Cultivars)

Science base

The breeder usually oversees the entire process, therefore acts more like a process engineer than a scientist

The requirements from plant breeders

- Plant breeding is about making lots of decisions in a limited time and based on a limited information

- The new development is providing the breeder with new tools
  - understanding genotype, quantitative traits - in the past;
  - genome sequences - today.

- The breeder will have more information to process in the same amount of time, e.g. 0-30 days between harvest in seeding of the next crop (season or counter-season).
The requirements from plant breeders
- The breeder will need a system to simplify the information processing and decision making
- Similar changes happened before, e.g. plot combine dramatically increased the size of breeding programs.
- The new information will not completely replace the old tools, and definitely not field testing.
- The linkages break, most models in genome selection are valid only for 2-3 generations of recombination.
- A constant fine-tuning of the model/system will be required from teams of breeders, molecular biologists, biostatisticians pathologists etc...
- GxE interaction will still be there to guarantee the job security for the breeders.

Exercise

- Goals for your crop
- How to prioritize?
1.1.3. Germplasm resources and pools

Germplasm collection and preservation

**Guaranteed food security: two major challenges**

- Increasing world population
- Climate change & crop yield
Increase yield, but not only….

- One billion people suffer from hunger nowadays.
- The world can feed using the agricultural area available ca. 20 billion people.
- Combatting hunger is also a matter of sound economical/political management.

Loss of biodiversity: genetic erosion

- What do we need to develop improved varieties?
  - most importantly access to a broad spectrum of biodiversity (genetic variation)
- But….
  - loss of biodiversity is taking place to a large extent
- Consequently: there is a clear need to halt this loss of biodiversity
Protection of biodiversity

- **In situ** (nature reserves, on farm management)
- **Ex situ** (genebanks, botanical gardens)

**Genebanks**
- purpose: to collect, conserve and make genetic resources available for breeding and research
- majority of genebanks established since 1970s
- currently ca. 150 genebanks manage ca. 7.4 million accessions (ca. 1.5 – 2 million unique)
Germplasm resources management system

– International centres

– National centres

– Crop (or group of crops) specific
  • Collection and cataloguing
  • Preservation and Regeneration
  • Rarely characterization
International

- CYMMYT International Maize and Wheat Improvement Center (Centro, International de Mejoramiento de Maiz y Trigo) – El Batán, Ciudad Mexico, Mexico
  - Wheat, Maize (8000)

- IRRI – International Rice Research Institute – Los Banos, Philippines
  - Rice (42000)

- ICRISTAT – International Crops Research Institute for Semi-Arid Tropics, Hyderabad, India
  - Sorghum

- CIP – International Potato Center, Lima, Peru
  - Potato (11000)

International

- IPGRI – International Plant Genetic Resources Research Institute – Rome, Italy
  - Germplasm conservation (world agency)

- ICARDA – International Center for Agricultural Research for Dry Areas, Aleppo, Syria
  - Barley, Wheat, Lentils

- CIAT – International Center For Tropical Agriculture, Palmira, Colombia
  - Cassava, Bean (10000)

- IITA – International Institute for Tropical Agriculture, Ibadan, Nigeria
  - Legumes, tubers
### International

- **N.I. Vavilov Institute of Plant Industry (Всероссийский Институт Растениеводства им. Н.И.Вавилова)** – St. Petersburg, Russia  
  - Plant diversity - all crops (320,000)

- **WARDA – West African Rice Development Association, Monrovia, Liberia**  
  - Rice

- **USDA-ARS, Beltsville, USA**

### US National Plant Germplasm System

Germplasm maintained at 4 regional centers and 21 special centers

**Regional Centers:**

- **Ames, IA (NC - maize, sugarbeet, Cucumis)**

- **Geneva, NY (NE - pea, tomato)**

- **Griffin, GA (S - peanut, pepper, sorghum)**

- **Pullman, WA (W - alfalfa, beans)**
Special Centers:

- Aberdeen, ID (barley, oats, rice, rye, wheat)
- Brookings, SD (native grasses)
- Brownwood, TX (chestnut, hickory, pecan)
- Byron, GA (bamboo)
- College Station, TX (cotton)
- Columbia, MO (wheat genes)

Special Centers:
- Corvallis, OR (filbert, hop, mint, pear, small fruit)
- Davis, CA (stone fruit, figs, grapes)
- Fargo, ND (flax)
- Hilo, H (tropical fruits)
- Lexington, KY (clover)
- Madison, WI (crucifer genes)
- Miami/Mayaguez, FL/PR (tropical crops)
Special Centers:
• Miami/Mayaguez, FL/PR (tropical crops)
• Orlando/Leesburg, FL (citrus)
• Oxford, NC (tobacco)
• Raleigh, NC (gama grass)
• Riverside/Brawley, CA (citrus, date)

Special Centers:
• Salinas, CA (lettuce)
• Stillwater, OK (wild peanut)
• Stoneville, MS (southern soybean)
• Sturgeon Bay, WI (potato)
• Urbana, IL (northern soybean)
• **US Organizations**
  - American Public Gardens Association (APGA), Delaware
  - Center For Plant Conservation (CPC), St. Louis, Missouri
  - Cereal Crops Research Unit, Madison, Wisconsin
  - Maize Genetics and Genomics Database (MaizeGDB)
  - North American Plants Collections Consortium (NAPCC)
  - California Rare Fruit Growers, Inc.
  - Forage Information system
  - Germplasm Enhancement of Maize (GEM)
  - Grains Genome Database (GrainGenes)
  - GRAMENE: A Resource for Comparative Grass Genomics
  - National Research Support Project-5
  - Northwest Berry and Grape Infonet, Oregon
  - Plant Conservation Alliance
  - Plants Database, USDA, Natural Resources Conservation Service
  - Seeds of Success
  - Solanaceae Genome Database
  - Soybean Genome Database (SoyBase)
  - US Cultivars Eligible for OECD Certification (OECD data maintained by ARS)
  - Western Gulf Forest Tree Improvement Program
### International Organizations

- **Arable crop genebank and online database**, Christchurch, New Zealand
- **Asian Vegetable Research and Development Center** (AVRDC), Shanhua, Taiwan
- **Bioprospecting International (formerly IPGRI)**, Rome, Italy
- **Tropical Agricultural Research and Higher Education Center** (CATIE), Costa Rica
- **Centre for Genetic Resources**, The Netherlands (CGN)
- **Centre for Plant Genetic Resources Poland**, Blonie, Poland
- **Centro Internacional de Mejoramiento de Maíz y Trigo** (CIMMYT), Mexico
- **Chinese Crop Germplasm Resources**, Beijing, China
- **CGIAR's System-wide Information Network for Genetic Resources** (SINGER)
- **Conservation, Evaluation, Exploitation and Collection of Minor Fruit Tree Species**, European Community Project
- **The Consultative Group on International Agricultural Research** (CGIAR)
- **Medicinal Plants Conservation Project of Egypt** (MPCP), Cairo, Egypt
- **European Union's Regulation 1467/94, Genetic Resources in Agriculture**
- **European Vitis Database**
- **Desert Research Center**, North Sinai Desert Station, Sinai, Egypt

---

### International Organizations

- **The Greek Vitis Database**, Heraklion, Crete, Greece
- **Instituto Nacional de Biodiversidad** (INBio), Costa Rica
- **International Center for Agricultural Research in the Dry Areas** (ICARDA), Aleppo, Syria
- **International Center for Tropical Agriculture**, (CIAT) Cali, Colombia
- **International Crops Research Institute for the Semi-Arid Tropics**, Patancheru, India
- **International Institute of Tropical Agriculture** (IITA), Ibadan, Nigeria
- **International Potato Center** (CIP), Lima, Peru
- **International Rice Research Institute** (IRRI), Los Banos, Philippines
- **International Registration Authorities** (IRAs)
- **International Sorghum and Millet (INTSORMIL) Collaborative Research Support Program** (CRSP), Lincoln, Nebraska
- **Istituto di Genetica Vegetale (Germplasm Research Institute)**, Bari, Italy
- **Ministry of Agriculture, Forestry and Fisheries Genebank**, Japan
- **National Bureau of Plant Genetic Resources** (NBPRG), New Delhi, India
International Organizations

- National Center for Genetic Resources, and Biotechnology Research (CENARGEN) of the Brazilian Organization for Agricultural Research (EMBRAPA), Brasil
- National Plant Genetic Resources Center, Taiwan
- National Research Centre, (INTA) Castelar, Argentina
- The Nordic Gene Bank, Alnarp, Sweden
- Plant Gene Resources of Canada, Saskatoon, Canada
- Plants Genetic Resources in Central Asia and Caucasus
- PROSEA, Plant Resources of South-East Asia
- The Research Institute of Crop Production, Czech Republic
- Rice Genetic Resources in Japan
- Royal Botanic Gardens, Kew, UK
- Shared Information of Genetic Resources SHIGEN, Japan
- UK Plant Genetic Resources Group
- N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR), St. Petersburg, Russia
- Integrated Wheat Science Database, Japan
- Information Centre for Genetic Resources (IGR) (ZADI)

Example: Centre for Genetic Resources the Netherlands
Collecting expeditions

• Since 1955
  – 26 expeditions (21 abroad & 5 NL); 18 since est. CGN in 1985
  – ca. 5000 accessions sampled -> (ca. 2100 in CGN genebank; 8% of CGN genebank content)

• Current expeditions abroad
  – targets (species, areas) determined in cooperation with breeding companies; expeditions are financed by breeding companies and Min. of Agriculture
Recent collecting expeditions; 2009 leek CWR Greece

Why a CGN leek expedition 2009?

- Based on CGN’s gap analysis of leek and on request of leek breeding companies
- Global *Allium porrum* and CWR genetic resources:

<table>
<thead>
<tr>
<th></th>
<th>GRIN</th>
<th>SINGER</th>
<th>EURISCO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. ampeloprasum</em></td>
<td>49</td>
<td>2</td>
<td>47</td>
<td>98</td>
</tr>
<tr>
<td><em>A. bourgeau</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>A. commutatum</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>A. porrum</em></td>
<td>51</td>
<td>12</td>
<td>1005</td>
<td>1068</td>
</tr>
</tbody>
</table>
Collecting route

Result of the CGN leek expedition 2009

<table>
<thead>
<tr>
<th></th>
<th>GRIN</th>
<th>SINGER</th>
<th>EURISCO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ampeloprasum</td>
<td>49</td>
<td>2</td>
<td>47 + 60</td>
<td>98 + 60</td>
</tr>
<tr>
<td>A. bourgeai</td>
<td>0</td>
<td>0</td>
<td>1 + 20</td>
<td>1 + 20</td>
</tr>
<tr>
<td>A. commutatum</td>
<td>0</td>
<td>0</td>
<td>0 + 20</td>
<td>0 + 20</td>
</tr>
<tr>
<td>A. porrum</td>
<td>51</td>
<td>12</td>
<td>1005</td>
<td>1068</td>
</tr>
</tbody>
</table>
Svalbard Global Seed Vault (Global Crop Diversity Trust; Nordic Gene Bank)

- On the Norwegian island of Spitsbergen in the remote Svalbard archipelago; 1300 km (810 mi) from the North Pole
  - 491058 accessions
  - 3487 species
  - 224 countries

Seed storage

- Storage conditions:
  - -20 C
  - 3-7% moisture content
- Minimum germination requirements for seed storage:
  - cvs and landraces: 80%
  - wild species: 60%
- Monitoring germination
- Duplication of collections
Incorporation of genetic variability from different sources into a commercial breeding program

• Cultivars
  – Old
  – Competitors
    • Intellectual property issues
  – Non-target markets (non-adapted, exotic)

• Landraces
• Wild relatives

How do we incorporate germplasm resources into a breeding program?

**Gene Pools Strategy**
(not to be confused with gene pools in heterotic patterns)

• Plant Breeding Strategy
  – Breeder should develop a strategy with primary and secondary gene pools (breeding, pre-breeding)
    – Gene flow: Secondary pool $\rightarrow$ Primary pool $\rightarrow$ Cultivars
    – May have several populations in each pool
Gene Pools Strategy

• Primary gene pools
  – Definition: populations made up by crossing elite germplasm (own or competitors’)
  – Objective: develop and release cultivars (breeding)
  – Characteristics: high mean, low variance for traits of interest
  – Uses: many small population with useful combinations

Gene Pools Strategy

• Secondary gene pools
  – Definition: populations made up by crossing with exotic and/or wild germplasm
  – Objective: identify new traits, improve germplasm adaptation (pre-breeding)
  – Characteristics: low mean, high variance for traits of interest
  – Uses: few, large populations having traits of possible future interest
Germplasm collection screening priority

Primary pool
- Adapted cultivars
- Breeding lines
- Obsolete cultivars

Secondary pool
- Core collections
- All plant introduction accessions
- Wild and exotic germplasm
- Related species
- Mutagen-treated populations

Class discussion
- Do we have good access to genetic variation in our crops/markets?

- What is the major barrier to accessing the genetic variation?

- Do you see any of these in your crop/geography situation:
  - Lack of genetic diversity
  - Intellectual property rights
  - Lack of knowledge of how to incorporate exotic material into our commercial breeding program

- Have you used a gene bank to access genetic material?
1.2.3 Inbreeding Coefficient

**Inbreeding coefficient**

- Inbreeding results when two related individuals are mated.
- Two individuals are related if they have at least one ancestor in common.
- $2^t$ ancestors ($t =$ number of generations)
- Relatedness is only meaningful if it does not go beyond a specific point of time.
Two types of identity of alleles, or differentiation based on

- Functionality of alleles, or physical state - **alike in state** (physically represent the same allele)
- origin of alleles – **identical by descent** (present in a common ancestor)

- If two alleles are identical by descent, the alleles are **autozygous**;
- if not, they are **allozygous**.

The coefficient of inbreeding $F$

- the probability that, at a single locus, the two alleles in the same individual are identical by descent

- $F=0$ indicates no inbreeding

- $F=1$ indicates complete inbreeding
The coefficient of co-ancestry

- between an individual X and an individual Y is the probability that, at a single locus, a random allele from X and a random allele from Y are identical by descent.

- \( f_{XY} = 0 \) indicates an unrelated relationship (for the given locus)

- \( f_{XY} = 1 \) indicates that the individuals are homozygous for the copies of same alleles found in a common ancestor

- \( f_{XY} = 1 \) across all loci indicates that the individuals are fully inbred and genetically identical

- The coefficient of inbreeding measures the identity by descent within an individual

- The coefficient of co-ancestry measures the identity by descent between two individuals
  - (Or within an individual in self-pollinating crops)

- \( F \) of an individual in progeny from two parents X and Y is equal to \( f_{XY} \)
Calculating the coefficient of inbreeding - Probability of identity by descent

\[ P(A_1A_1) = \left(\frac{1}{2}\right)^4 = \frac{1}{16} \]

\[ P(A_2A_2) = \left(\frac{1}{2}\right)^4 = \frac{1}{16} \]

\[ P(A_1A_1) \text{ or } (A_2A_2) = 2\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \]

If parent A was an inbred with \( F_A \), then:

\[ F_X = \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^3 F_A = \left(\frac{1}{2}\right)^3 \left(1 + F_A\right) \]

Number of individuals in the path connecting the common ancestors allows the calculation of \( F \).

\[ F_X = \Sigma \left(\frac{1}{2}\right)^n \left(1 + F_A\right) \]

When multiple paths can be constructed, each path contributes additional probability. \( F \) is then calculated as the sum of all individual probabilities for paths through which parents are related.
Exercise
Calculate the inbreeding coefficient for the following pedigrees

Example 5.1 from page 84 in Falconer and Mackay

Example with # individuals in the path – (Falconer & Mackay, page 83 5.2) Connect C and D

Calculation of Inbreeding Coefficient

\( F \) defined as increase of homozygosity

- Alleles that are identical by descent within an individual are always in a homozygous form.

- \( F \) coefficient can also be defined based on increase of homozygocity due to identity by descent

- \( F \) coefficient is equal to the proportion by which heterozygocity is reduced upon inbreeding, relative to a population in a HW equilibrium.

- If \( P_{12}(H) \) is the frequency of heterozygotes at equilibrium and \( P_{12}(F) \) is frequency of heterozygotes upon inbreeding, the \( F \) coefficient is equal to

\[
F = 1 - \frac{P_{12}(F)}{P_{12}(H)}
\]
Example

- Selfing in the $F_2$ leads to genotype frequencies of
  - 37.5% $A_1A_1$; 25% $A_1A_2$; 37.5% $A_2A_2$ in $F_3$.

The coefficient of inbreeding among $F_3$ plants is

\[
P_{12}^{(F)}
\]

\[
F = 1 - \frac{P_{12}}{0.5} = 0.25/0.50 = 0.5
\]

- The increase in the coefficient of inbreeding is halved upon each additional generation of selfing. In the $n$th generation, the $F_n$ coefficient equals

\[
F_n = 1 - \left(\frac{1}{2}\right)^{n-2}
\]

<table>
<thead>
<tr>
<th>Generation of:</th>
<th>Frequency of heterozygous plants of families</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plants $F_2$ of $S_0$</td>
<td>$F_3$ or $S_1$</td>
<td>$P_{12}$</td>
</tr>
<tr>
<td>$F_3$ or $S_1$</td>
<td>$F_4$ or $S_2$</td>
<td>$(0.5)P_{12}$</td>
</tr>
<tr>
<td>$F_4$ or $S_2$</td>
<td>$F_5$ or $S_3$</td>
<td>$(0.25)P_{12}$</td>
</tr>
<tr>
<td>$F_5$ or $S_3$</td>
<td>$F_6$ or $S_4$</td>
<td>$(0.125)P_{12}$</td>
</tr>
<tr>
<td>$F_n$ or $S_{n-2}$</td>
<td>$F_{n+1}$ or $S_{n-1}$</td>
<td>$(1/2)^{n-2}P_{12}$</td>
</tr>
<tr>
<td>$F_\infty$ or $S_\infty$</td>
<td>$F_\infty$ or $S_\infty$</td>
<td>0</td>
</tr>
</tbody>
</table>

Corn example—others $F_2$ =% heterozygosity =50%
Summary
1. $F_Z$ of progeny = $f_{XY}$ of parents
2. Co-ancestry = of average co-ancestry of previous generation
3. Calculate the number of individuals in the path
4. Add all possible paths

Summary
When relatives are mated, gametes can no longer be considered as randomly sampled.

1. Relatives share more alleles in common than non-relatives.

2. The offspring of relatives have a higher probability of having shared alleles and shared genotypes than offspring from randomly mated parents.
Summary

Primary effects
- Most genotypes in highly inbred populations will be homozygous; the consequences of this excess homozygosity can be beneficial or detrimental.

- Individual lines chosen from inbred populations will breed true and remain genetically uniform; inbred populations can be heterogeneous.

Important implications
1. Inbreeding affects the efficiency of selection.
2. Inbreeding restricts recombination
Inbreeding - Inbreeding coefficient


Introduction

Inbreeding results when two related individuals are mated. Two individuals are related if they have at least one ancestor in common.

Consideration should be given to the number of generations

In a population of bisexual organisms every individual has

- 2 parents
- 4 grandparents
- 8 great-grandparents etc...

$t$ generations back it has $2^t$

- 30 generations ~ 1 billion
- 40 generation ~ 1 trillion

Any pair of individuals therefore must be related in a more or less remote past. Therefore the relatedness only becomes meaningful if we specify a point in time beyond which the ancestors will not be a factor.

- Consideration should be given to the population size
- Pairs mating at random are closely related in a smaller population than in a large one

A distinction must be made among the alleles according to their origin

- $A_1$ allele in one plant is treated as $A_1$ allele in any other plant

Two types of identity of alleles, or differentiation based on

- Functionality of alleles, or physical state - alike in state (physically represent the same allele)
- origin of alleles – identical by descent (present in a common ancestor)
- If two alleles are identical by descent, the alleles are autozygous; if not, they are allozygous.

I-1 individual is mated with an individual that carry $A_2$ allele

Assume that two $A_1$ alleles are not identical by decent

Individual II-1 and II-2 receive the same copy of the allele $A_1$; they are identical by state and by descent

Individual II-3 receives a copy of the other $A_1$; identical by state but not but descent
Individual III-1 is an inbred; individual III-2 is not an inbred

Whether or not two \( A_1 \) alleles are identical by descent cannot be known unless some methods are used for tagging each copy of \( A_1 \) in I-1 and tracing each copy throughout the pedigree.
However, the probability that two alleles are identical by descent can be deducted from:
- the mating system
- pedigree structure

The coefficient of co-ancestry (also known as coefficient of parentage) between an individual X and an individual Y is the probability that, at a single locus, a random allele from X and a random allele from Y are identical by descent.
- \( f_{XY}=0 \) indicates no relationship (for the given locus)
- \( f_{XY}=1 \) indicates that the individuals are homozygous for the copies of same alleles found in a common ancestor
- \( f_{XY}=1 \) across all loci indicates that the individuals are fully inbred and genetically identical

The coefficient of inbreeding \( F \) could be defined in two ways:
- the probability that, at a single locus, the two alleles in the same individual are identical by descent
- \( F=0 \) indicates no inbreeding
- \( F=1 \) indicates complete inbreeding
- The coefficient of inbreeding measures the identity by descent within an individual
- The coefficient of co-ancestry measures the identity by descent either within an individual or between two individuals
- \( F \) of an individual in progeny from two parents X and Y is equal to \( f_{XY} \)

**Calculating the coefficient of inbreeding \( F \)**

\( F \) defined as increase of homozygocity

Alleles that are identical by descent within an individual are always in a homozygous form.
\( F \) coefficient can also be defined based on increase of homozygocity due to identity by descent
\( F \) coefficient is equal to the proportion by which heterozygocity is reduced upon inbreeding, relative to a population in a HW equilibrium.
If \( P_{12} \) is the frequency of heterozygotes at equilibrium and \( P_{12(f)} \) is frequency of heterozygotes upon inbreeding, the \( F \) coefficient is equal to
\[
P_{12(f)}
\]
\[
F = 1 - \frac{P_{12(f)}}{P_{12}}
\]
Lecture 1.2.3  

Inbreeding Coefficient  

Van Deynze

The expected frequency of heterozygoetes with arbitrary allele frequencies and values of \( F \) can be determined by rearranging the above equation. The frequency of heterozygotes at HW equilibrium is equal to \( P_{12} = 2pq \); therefore

\[
P_{12(F)} = 2pq(1-F)
\]

Inbreeding decreases the frequency of heterozygotes by \( P_{12} - P_{12(F)} = 2pqF \)

The frequency of \( A_1 A_1 \) proportionately increase to

\[
P_{11(F)} = p^2 + pqF
\]

and the frequency of \( A_2 A_2 \) proportionately increases to

\[
P_{22(F)} = q^2 + pqF
\]

Note that the inbreeding itself does not change the allele frequencies, but only genotype frequencies.

Therefore: mating system + population structure would allow us to deduct the coefficient of inbreeding.

Examples: Selfing

An \( F_2 \) population with genotype frequencies of 25% \( A_1A_1 \); 50% \( A_1A_2 \); 25% \( A_1A_2 \)

The \( F_2 \) population by virtue of its having the expected frequencies at HW equilibrium is equivalent to \( S_0 \) generation (prior to selfing). The \( F_2 \) generation is therefore considered a non-inbred generation with \( F=0 \), and it is used as a reference to interpret the \( F \) coefficient in the subsequent selfing generations.

The degree of relationship expressed as inbreeding coefficient is essentially a comparison between the population in question and a specified or implied population used as a point of reference.

Selfing in the \( F_2 \) leads to genotype frequencies of 37.5% \( A_1A_1 \); 25% \( A_1A_2 \); 37.5% \( A_1A_2 \) in \( F_3 \).

The coefficient of inbreeding among \( F_3 \) plants is

\[
F = 1 - \frac{P_{12(F)}}{P_{12}} = 1 - 0.25/0.50 = 0.5
\]

The increase in the coefficient of inbreeding is halved upon each additional generation of selfing (table bellow). In the \( n^{th} \) generation, the \( F_n \) coefficient equals

\[
F_n = 1 - (1/2)^{n-2}
\]

This equation assumes the \( F_2 \) population is non-inbred, which is true if the two inbred crossed to for the \( F_1 \) are unrelated (i.e. \( f_{XY}=0 \)). In breeding programs \( F_2 \) is assumed non-inbred even if the parental lines are related. This assumption allows that genetic parameters such as the population mean and variance are defined and estimated for non-inbred based population.

The actual level of inbreeding is greater with related parents than with unrelated.

58
<table>
<thead>
<tr>
<th>Generation of:</th>
<th>Frequency of heterozygous plants or families</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_2$ of $S_0$</td>
<td>$F_3$ or $S_1$</td>
<td>$P_{12}$</td>
</tr>
<tr>
<td>$F_3$ or $S_1$</td>
<td>$F_4$ or $S_2$</td>
<td>$(0.5)P_{12}$</td>
</tr>
<tr>
<td>$F_4$ or $S_2$</td>
<td>$F_5$ or $S_3$</td>
<td>$(0.25)P_{12}$</td>
</tr>
<tr>
<td>$F_5$ or $S_3$</td>
<td>$F_6$ or $S_4$</td>
<td>$(0.125)P_{12}$</td>
</tr>
<tr>
<td>$F_n$ or $S_{n-2}$</td>
<td>$F_{n+1}$ or $S_{n-1}$</td>
<td>$(1/2)^{n-2}P_{12}$</td>
</tr>
<tr>
<td>$F_{\infty}$ or $S_{\infty}$</td>
<td>$F_{\infty}$ or $S_{\infty}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Deviation from idealized population

- Effective population size (number of breeding individuals) – smaller than the total
- Exclusion due to close related mating
  - Bisexual, self-incompatibility,
- Different number of males and females
- Different numbers in successive generations
Identity by descent in pedigrees

When the population does not have a well defined family structure, i.e. irregular pedigrees, the coefficient of inbreeding is easier to interpret when defining the probability of identity by descent.

Probability of identity by descent

![Pedigree diagram]

\[ P(A_1A_1) = \left(\frac{1}{2}\right)^4 = \frac{1}{16} \]
\[ P(A_2A_2) = \left(\frac{1}{2}\right)^4 = \frac{1}{16} \]
\[ P(A_1A_1) \text{ or } (A_2A_2) = 2\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \]

How does the inbreeding coefficient of A contribute?

How do inbreeding coefficients of B and C contribute?

\[ F_X = \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^3 F_A = \left(\frac{1}{2}\right)^3 (1 + F_A); \text{ Inbreeding coefficient of individual} \]

Number of individuals in the path connecting the common ancestors allows the calculation of \( F \).

Example with 6 individuals in the path – (Falconer & Mackay, page 83)

When multiple paths can be constructed, each path contributes additional probability. \( F \) is then calculated as the sum of all individual probabilities for paths through which parents are related.

\[ F_X = \sum \left(\frac{1}{2}\right)^n (1 + F_A) \]
Co-ancestry or kinship

The co-ancestry of P and Q is completely determined by the co-ancestry of A, B, C and D, and it can be shown that it is the mean of four possible co- ancestries AC, AD, BC, BD.

When expressed in terms of probabilities:
- Probability that \( p_1 \) would be from A is \( \frac{1}{2} \) (0.5)
- Probability that \( q_1 \) would be from C is \( \frac{1}{2} \) (0.5)
- The probability of both is \( 0.5 \times 0.5 = 0.25 \) (¼)

Repeat this for AD, BC and BD

\[ f_{PQ} = \frac{1}{4} f_{AC} + \frac{1}{4} f_{AD} + \frac{1}{4} f_{BC} + \frac{1}{4} f_{BD} \]

or

\[ F_X = f_{PQ} = \frac{1}{4} (f_{AC} + f_{AD} + f_{BC} + f_{BD}) \]
Co-ancestries

**Individual to itself**

\[ f_{AA} = \frac{1}{2} (1 + F_A) \]

\[ f_{AA} = 0.5 \text{ if } F_A = 0 \text{ (non-inbred); } f_{AA} = 1 \text{ if } F_A = 1 \]

**Offspring to its parents**

\[
\begin{array}{c}
A \\
\downarrow \\
P \\
\downarrow \\
X \\
\uparrow \\
C \\
\downarrow \\
Q \\
\downarrow \\
D \\
\uparrow \\
B
\end{array}
\]

E.g. co-ancestry of P with A

\[ f_{AP} = \frac{1}{4} (f_{AB} + f_{AA}) \]

i.e. the average of co-ancestries of P’s parents (AB) and A.

If A and B not related; and A non-inbred; then

\[ f_{AB} = 0.0; \quad f_{AA} = 0.5; \quad \text{and } f_{AP} = 0.25 \]

If inbred parents \( f_{AB} = 0; \quad f_{AA} = 1; \quad \text{and } f_{AP} = 0.25 \)

**Full sibs**

\[
\begin{array}{c}
A \\
\downarrow \\
P \\
\downarrow \\
X \\
\uparrow \\
C \\
\downarrow \\
Q \\
\downarrow \\
B
\end{array}
\]

\[ f_{PQ} = \frac{1}{4} (2f_{AB} + f_{AA} + f_{BB}) \]

i.e. the average of co-ancestries of P and Q (not that \( f_{AB} = f_{BA} \))

with no previous inbreeding: \( f_{PQ} = \frac{1}{4} (2*0 + 0.5 + 0.5) = \frac{1}{4} \)

If parents are inbreds, then \( f_{PQ} = \frac{1}{4} (2f_{AB} + f_{AA} + f_{BB}) = \frac{1}{4} (2f_{AB} + 1 + 1) = \frac{1}{4} * 2 (1 + f_{AB}) = \frac{1}{2} (1 + f_{AB}) \)
Half sibs

\[
\begin{array}{c}
B \\
\downarrow \\
P \\
\downarrow \\
X \\
\downarrow \\
C \\
\downarrow \\
Q
\end{array}
\]

Already demonstrated that \( f_{PQ} = 1/8 \)

Summary
When relatives are mated, gametes can no longer be considered as randomly sampled.
1. Relatives share more alleles in common than non-relatives.
2. The offspring of relatives have a higher probability of having shared alleles and shared genotypes than offspring from randomly mated parents.

Primary effects
- Most genotypes in highly inbred populations will be homozygous; the consequences of this excess homozygosity can be beneficial or detrimental.
- Individual lines chosen from inbred populations will breed true and remain genetically uniform; inbred populations can be heterogeneous.

Important implications
1. Inbreeding affects the efficiency of selection.
2. Inbreeding restricts recombination.

Exercise 1.2.4.1.
Calculate the inbreeding coefficient for the following pedigrees

a. Example 5.1 from page 84 in Falconer and Mackay (open book)

b. Malting barley MN77-825 (University of Minnesota, Rasmusson and Phillips)

Discussion:
The importance of understanding the inbreeding in different crops: cross pollinating, self-pollinating, hybrid crops.
### General formula

- \( f_{PQ} = \frac{1}{4} (f_{AC} + f_{AD} + f_{BC} + f_{BD}) \)

### Overlapping generations

- \( f_{PC} = \frac{1}{2} (f_{AC} + f_{BC}) \)
- \( f_{PD} = \frac{1}{2} (f_{AD} + f_{BD}) \)
- \( f_{PQ} = \frac{1}{2} (f_{PC} + f_{PD}) \)

### Coancestries

#### Self

- \( f_{AA} = \frac{1}{2} (1 + F_A) \)
- \( F_{parent} = 0 \quad F_{parent} = 1 \)

#### Offspring and parent

- \( \frac{1}{2} (f_{AB} + f_{AA}) \)

### Full-sibs

- \( \frac{1}{4} (2f_{AB} + f_{AA} + f_{BB}) \)

### Half-sibs

- \( \frac{1}{4} (f_{AB} + f_{AC} + f_{BC} + f_{AA}) \)
Regular system of inbreeding (same mating system in every generation)
*note that \( t \) = generation of interest; \( t-1 \) the previous; \( t-2 \) the one before

**Self-fertilization, recurrent formula**
\[
F_X = f_{AA} = 1/2(1+F_A) \\
F_t = 1/2(1+F_{t-1})
\]

**Full-sibs, recurrent formula**
\[
F_X = f_{PQ} = 1/4(2f_{AB} + f_{AA} + f_{BB}) \\
F_t = 1/4(1+2F_{t-1}+F_{t-2})
\]

**Offspring-parent, recurrent formula**
\[
F_X = f_{PA} = 1/2(f_{AB} + f_{AA})
\]
recurrent formula same as full-sibs

**Half-sibs, recurrent formula**
\[
F_t = 1/8(1+6F_{t-1}+F_{t-2})
\]
or when B and C always full sibs
\[
F_t = 1/16(3+8F_{t-1}+4F_{t-2}+F_{t-3})
\]

\( B \times A \)  
Repeated backcrossing, recurrent formula
\[
F_X = f_{AD} = 1/2(f_{AA} + f_{AC} = 1/2[1/2(1+F_A)+F_D])
\]

\( C \times A \)
\[
F_t = 1/4(1+F_A + 2F_{t-1})
\]

\( D \times A \)
\[
F_t = 1/4(1+F_A + 2F_{t-1})
\]

where \( F_A \) of the recurrent parent
Quantitative genetic model

Intro

1.2.4

Quantitative genetic model

1. Introduce the concept of values
   a. Phenotypic and genotypic values
   b. Learn to calculate the population mean
   c. Multiple loci extension
   d. Calculate the progeny mean from parent genotypes

2. Average effect of alleles

3. Breeding values and dominance deviation

4. Phenotypic and genotypic variances

5. Additive and dominance genetic variances
   - Now we know what they are but not how to calculate them

6. Covariance between relatives common in plant breeding
• Cross between two parents $A_1A_1 \times A_2A_2$
• $F_2$ range 100-200, mean 150,
• $BC_1$ range 75-125, mean 100

• Allele frequency in $F_2$, $BC_1$
• Do they explain the differences alone?

---

**Quantitative genetic model**

**Introduction – Values**

Phenotypic value for individual with genotype $A_iA_j$

\[ P_{(ij)k} = G_{ij} + e_{(ij)k} \]

....or.... \[ P_{(ij)k} = \mu + g_{ij} + e_{(ij)k} \]

$G_{ij} = \mu + g_{ij}$; the genotypic value = population mean plus the deviation due to the $A_iA_j$ genotype
Population mean for One-Locus model

- a single locus A with $A_1$ and $A_2$ are alleles, $f(A_1) = p$:

\[
\begin{array}{cccc}
A_2A_2 & A_1A_2 & A_1A_1 \\
-a & 0 & a
\end{array}
\]

- The point of zero value = mid point between the two homozygotes.
- $d$ – the value of the heterozygote
- $d = 0 \rightarrow$ no dominance
- $d = a \rightarrow$ complete dominance
- $d > a \rightarrow$ overdominance

Example: Consider Mendel’s tall and dwarf pea plants.

- i. Plant height is controlled by a single dominant locus, dominant phenotypes (TT and Tt) are 3’ tall, the recessive genotype (tt) is 1’ tall.
- ii. Determine the midpoint between alternate homozygotes
  - $(\text{value}(A_1A_1)\text{-value}(A_2A_2) = 2)$
- iii. Scale genotypic values by subtracting this midpoint; to unscale, add back this midpoint.

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>$A_2A_2$</th>
<th>$A_1A_2$</th>
<th>$A_1A_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original values</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Scaled values</td>
<td>$1-2 = -1 = -a$</td>
<td>$3-2 = 1 = d$</td>
<td>$3-2 = 1 = a$</td>
</tr>
</tbody>
</table>

iv. $a=1$, $d=1$
### Population mean

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Value</th>
<th>(F^*V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1A_1)</td>
<td>(p^2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_1A_2)</td>
<td>(2pq)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_2A_2)</td>
<td>(q^2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Population mean

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Value</th>
<th>(F^*V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1A_1)</td>
<td>(p^2)</td>
<td>(a)</td>
<td></td>
</tr>
<tr>
<td>(A_1A_2)</td>
<td>(2pq)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>(A_2A_2)</td>
<td>(q^2)</td>
<td>(-a)</td>
<td></td>
</tr>
</tbody>
</table>
**Population mean**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Value</th>
<th>( F^*V )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( p^2 )</td>
<td>( a )</td>
<td>( p^2a )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( 2pq )</td>
<td>( d )</td>
<td>( 2pqd )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>( q^2 )</td>
<td>(-a)</td>
<td>( q^2-a )</td>
</tr>
</tbody>
</table>

Population mean = \( \Sigma F^*V = p^2a + 2pqd - q^2a = a(p-q) + 2pqd \)

---

**Exercise – closed binders**

a. Mendel’s peas, midpoint = 2, \( a=d=1 \); calculate for \( p=q=0.5 \) vs \( p=0.8, q=0.2 \)

| \( q=f_{[A_2A_2]} \) | \( M=a(p-q) + 2pqd = \) Scaled mean Unscaled mean |
|-----------------|-----------------|-----------------|
| 0.5             |                 |                 |
| 0.2             |                 |                 |

b. At a segregating locus, the mean of an \( F_2 \) population between two inbred lines (\( p=q=0.5 \)):

c. At a segregating locus, the mean of a \( BC_1 \) to the superior parent (\( p=0.75; q=0.25 \)):

d. At a segregating locus, the mean of a \( BC_1 \) to the inferior parent (\( p=0.25; q=0.75 \)): 
Exercise

a. Mendel’s peas, midpoint = 2, a=d=1; calculate for p=q=0.5 vs p=0.8, q=0.2

<table>
<thead>
<tr>
<th>q=f(A_2A_2)</th>
<th>M= a(p-q) + 2pqd</th>
<th>Scaled mean</th>
<th>Unscaled mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1(0.5-0.5)+2(0.5)(0.5)</td>
<td>1</td>
<td>2.50</td>
</tr>
<tr>
<td>0.2</td>
<td>1(0.8-0.2)+2(0.8)(0.2)</td>
<td>0.92</td>
<td>2.92</td>
</tr>
</tbody>
</table>

b. At a segregating locus, the mean of an F_2 population between two inbred lines (p=q=0.5):

c. At a segregating locus, the mean of a BC_1 to the superior parent (p=0.75; q=0.25):

d. At a segregating locus, the mean of a BC_1 to the inferior parent (p=0.25; q=0.75):

Extension to multiple loci

- Assumption: combination of allele contribution from different loci as simple addition (no epistasis), then:

\[ \mu = \Sigma a(p-q) + \Sigma 2pqd \]
**Extension to multiple loci**

Example closed binders

a. With four loci and all $a_i = d_i = 1$
   - Value for $A_1A_1 B_1B_2 C_1C_2 D_2D_2$

b. With four loci and all $a_i = 1$, $d_i = \frac{1}{2}$
   - Value for $A_1A_1 B_1B_2 C_1C_2 D_2D_2$
Quantitative genetic model

**Introduction - Values**

Genetic properties of a population have been established in terms of allele and genotype frequencies.

Example: A1A1 x A2A2 = F1, F2, BC1, elaborate on the allele frequencies, mean, range of values.

In order to establish the relationship between these and the quantitative differences expressed in a metric character, we need to introduce the concept of VALUE.

More specifically, we are interested in the value associated to a particular genotype. However, we cannot measure the genotypic value directly.

Phenotypic value = the value observed when a character is measured

To understand the genetic properties of a population, we need to partition the phenotypic value into components attributable to different causes (genotype and environment). For individual \( k \) with genotype \( A_iA_j \) the model is:

\[
P_{ij}k = G_{ij} + e_{(ij)k} \quad (+ G\times E): \text{Phenotypic value} = \text{Genotypic value} + \text{Environmental deviation} \quad (+ G \times E)
\]

An equivalent model is \( P_{ij}k = \mu + g_{ij} + e_{(ij)k} \)

Both models imply that phenotype is due to both genetic and nongenetic effect (*Genotypic Value* and *Environmental Deviation*).

It is safe to assume \( G_{ij} \) and \( e_{(ij)k} \) are uncorrelated.

From above \( G_{ij} = \mu + g_{ij} \) i.e. the genotypic value is the population mean plus the deviation due to the \( A_iA_j \) genotype (will follow up on this equation in the section on breeding value).

Note the choice of wording – \( G \) the value is conferred by the genotype of an individual; \( e \) is the environmental effect causing deviation from the value in one direction or other.

a. Therefore, the mean environmental deviation in the population as a whole is taken to be zero.

b. Consequently, the mean phenotypic value = to the mean genotypic value (of a population, of individual)
Population mean for One-Locus model

Consider a breeding population in a HW equilibrium at a single locus A with A1 and A2 are alleles, \( f(A_1) = p \). a, -a and d are the values assigned to the three genotypes i.e. genotypic values (convention is that A1 always increases the value)

\[
\begin{array}{c|c|c|c}
A_2A_2 & A_1A_2 & A_1A_1 \\
|----------|-------|----------|
-a       & 0      & d        & a
\end{array}
\]

- The origin or point of zero value is the mid point between the two homozygotes.
- d – the value of the heterozygote depends on the degree of dominance
  - \( d = 0 \) → no dominance
  - A1 dominant over A2 → d is positive
  - d = a (or –a) → complete dominance
  - d > a → overdominance
  - degree of dominance can be expressed as \( d/a \)

Example: Consider Mendel’s tall and dwarf pea plants.

i. Plant height is controlled by a single dominant locus, dominant phenotypes (TT and Tt) are 3’ tall, the recessive genotype (tt) is 1’ tall.

ii. Determine the midpoint between alternate homozygotes ( = value(A_1A_1)-value(A_2A_2) = 2)

iii. Scale genotypic values by subtracting this midpoint; to unscale, add back this midpoint.

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>A_2A_2</th>
<th>A_1A_2</th>
<th>A_1A_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original values</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Scaled values</td>
<td>1-2 = -1 = -a</td>
<td>3-2 = 1 = d</td>
<td>3-2 = 1 = a</td>
</tr>
</tbody>
</table>

iv. a=1, d=1

Population parameters – Mean

- Means using individual observations: \( \mu = \frac{\sum_{i=1}^{n} Y_i}{N} \)
- Means using frequencies: \( \mu = \frac{\sum_{i=1}^{n} f_i Y_i}{\sum f_i} \) (= \( \Sigma F^*V \))
Population mean

Allele $A_1$ occurs in frequency $p$ and $A_2$ in frequency $q$ ($=1-p$), with random mating,

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Value</th>
<th>$F*V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$p^2$</td>
<td>a</td>
<td>$p^2a$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$2pq$</td>
<td>d</td>
<td>$2pqd$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$q^2$</td>
<td>-a</td>
<td>$q^2-a$</td>
</tr>
</tbody>
</table>

population mean = $\mu = \Sigma F*V = p^2a+2pqd-q^2a = a(p-q) + 2pqd$ [because $p^2-q^2=(p+q)(p-q)=p-q$]

This is both the phenotypic and genotypic value of the population with respect to the character

The contribution of any locus to the population mean has two terms:

"$a(p-q)$" is the contribution of the homozygotes;

"$2pqd$" is the contribution of the heterozygotes.

When there is no dominance ($d=0$), the second term is $=0$

To unscale the population mean formula can be written as: $\mu = \overline{P} + a(p-q) + 2pqd$; ($\overline{P}$) midpoint value

the genotypic values, “$a$” and “$d$” are defined as deviation from the mean value of the two homozygotes (mid parent value).

The population mean (trait mean) changes as we change gene frequencies.

Exercises:

a. Mendel’s peas, midpoint = 2, $a=d=1$; calculate for $p=q=0.5$ vs $p=0.8$, $q=0.2$
Extension to multiple loci

If we assume combination of allele contribution from different loci as simple addition (no epistasis), then population mean is a sum of values attributable to the separate loci.

\[ \mu = \sum a(p-q) + \sum pqd \]  
(or \[ \mu = \sum P + \sum a(p-q) + \sum pqd \])

Examples

1. With additivity across loci, i.e. no epistasis, individual scaled genotypic values can be summed.

   a. With four loci and all \( a_i = d_i = 1 \)

      Value for \( A_1A_1 B_1B_2 C_1C_2 D_2D_2 \) = \( a_i + d_i + d_i + (-a_i) = 1 + 1 + 1 + (-1) = 2 \)

   b. With four loci and all \( a_i = 1, d_i = \frac{1}{2} \)

      Value for \( A_1A_1 B_1B_2 C_1C_2 D_2D_2 \) = \( a_i + d_i + d_i + (-a_i) = 1 + \frac{1}{2} + \frac{1}{2} + (-1) = 1 \)
### Quantitative genetic model

1.3.2 Variances

### Calculating the progeny mean from parent genotype

Progeny mean of progenies from parents of different genotype

<table>
<thead>
<tr>
<th>Offspring Genotypes (freq * value)</th>
<th>A₁A₁</th>
<th>A₁A₂</th>
<th>A₂A₂</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₁A₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₁A₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₂A₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Calculating the progeny mean from parent genotype**

**Progeny mean** of progenies from parents of different genotype

<table>
<thead>
<tr>
<th>Offspring Genotypes</th>
<th>Parent genotype</th>
<th>A1A1</th>
<th>A1A2</th>
<th>A2A2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1A1</td>
<td>p * a</td>
<td>q * d</td>
<td>0</td>
<td>p + q * d</td>
<td></td>
</tr>
<tr>
<td>A1A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2A2</td>
<td>0</td>
<td>p * d</td>
<td>-q a</td>
<td>p - q * a</td>
<td></td>
</tr>
</tbody>
</table>
Calculating the progeny mean from parent genotype

Progeny mean of progenies from parents of different genotype

<table>
<thead>
<tr>
<th>Offspring Genotypes</th>
<th>A_1A_1</th>
<th>A_1A_2</th>
<th>A_2A_2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent genotype</td>
<td>A_1A_1</td>
<td>A_1A_2</td>
<td>A_2A_2</td>
<td></td>
</tr>
<tr>
<td>p\ a</td>
<td>q\ d</td>
<td>0</td>
<td>p\ a + q\ d</td>
<td></td>
</tr>
<tr>
<td>(p/2)a</td>
<td>d/2</td>
<td>-(q/2)a</td>
<td>(p-q)a + d/2</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>p\ d</td>
<td>-q\ a</td>
<td>p\ d − q\ a</td>
<td></td>
</tr>
</tbody>
</table>

Example: single locus, a=1, d=1 (complete dominance for A1), p=q=0.5

Close Binder
Exercise: Multi-locus - two genotypes, \( a_i=1, d_i=1, p_i=q_i=0.5 \) for all loci: calculate the genotypic value and progeny mean

- \( G_1: A_1A_2, B_1B_2, C_1C_2 \)
- \( G_2: A_1A_1, B_1B_1, C_2C_2 \)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>( G_1 )</th>
<th>( G_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypic value</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Progeny mean</td>
<td>1.5</td>
<td>2</td>
</tr>
</tbody>
</table>

Compare the Genotypic values with Progeny means – How does this effect selection in F2?

### Average effect of alleles - intro

The average effect of an allele is: the average deviation from the population mean of individuals that received that allele from one parent; the other allele having came from the population at random.

<table>
<thead>
<tr>
<th>Type of gamete</th>
<th>Values and frequencies of genotypes produced</th>
<th>Mean value of genotypes produced</th>
<th>Population mean to be deducted</th>
<th>Average effect of an allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( a )</td>
<td>( a )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( d )</td>
<td>( -a )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( p )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( q )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Average effect of alleles**

The average effect of an allele is: the average deviation from the population mean of individuals that received that allele from one parent; the other allele having came from the population at random.

<table>
<thead>
<tr>
<th>Type of gamete</th>
<th>Values and frequencies of genotypes produced</th>
<th>Mean value of genotypes produced</th>
<th>Population mean to be deducted</th>
<th>Average effect of an allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(_1)A(_1)</td>
<td>A(_1)A(_2)</td>
<td>A(_2)A(_2)</td>
<td>a</td>
<td>d</td>
</tr>
<tr>
<td>A(_1)</td>
<td>p</td>
<td>q</td>
<td>pa + qd</td>
<td>[a(p-q) + 2dpq]</td>
</tr>
<tr>
<td>A(_2)</td>
<td>p</td>
<td>q</td>
<td>-qa + pd</td>
<td>[a(p-q) + 2dpq]</td>
</tr>
</tbody>
</table>
### Average effect of alleles

The average effect of an allele is: the average deviation from the population mean of individuals that received that allele from one parent; the other allele having came from the population at random.

<table>
<thead>
<tr>
<th>Type of Gamete</th>
<th>Values and Frequencies of Genotypes Produced</th>
<th>Mean Value of Genotypes Produced</th>
<th>Population Mean to be Deducted</th>
<th>Average Effect of an Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1A1</td>
<td>A2A2</td>
<td>a</td>
<td>pa + qd</td>
<td>a + d(q-p)</td>
</tr>
<tr>
<td>A1</td>
<td>A1A2</td>
<td>b</td>
<td>a(p-q) + 2dpq</td>
<td>a(qa + d(q-p))</td>
</tr>
<tr>
<td>A2</td>
<td>A2A1</td>
<td>c</td>
<td>-qa + pd</td>
<td>-p(a+ d(q-p))</td>
</tr>
</tbody>
</table>

Average effect of allele substitution is \( \alpha \)

### Average Effect of Allele Substitution

- Defined as the change in the mean of the offspring when the allele is changed to a different allele (e.g. \( A_2 \) changed to \( A_1 \)).

- \( \alpha = \alpha_1 - \alpha_2 = a + d(q-p) \)

- The average effect of alleles expressed in terms of average effect of allele substitution:
  
  \[
  \alpha_1 = qa \quad \text{and} \quad \alpha_2 = -p\alpha
  \]

*bb*
Breeding Value

- The value of an individual, judged by the mean value of its progeny, is called the *breeding value* of the individual.

- Breeding value can be measured and usually expressed as the deviation from the population mean.

- When mated with a number of individuals take at random from the population, breeding value is *twice* the mean deviation of the progeny from the population.

Breeding Value

- Defined in the terms of the average effects of alleles, the breeding value of an individual is sum of average effects of alleles that it carries.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Breeding Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>2α₁ = 2qα</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>α₁ + α₂ = (q-p)α</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>2α₂ = -2pα</td>
</tr>
</tbody>
</table>
The genotypic value of $A_iA_j$ can be decomposed from $G_{ij} = \mu + g_{ij}$ to $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$.

The genotypic value expressed as a deviation from the population mean is $g_{ij} = \alpha_i + \alpha_j + \delta_{ij}$.

The breeding value associated with $A_iA_j$ is defined as a sum of $\alpha_i$ and $\alpha_j$, sum of effects of individual alleles.

The term $\delta_{ij}$ is the dominance deviation associated with the $A_iA_j$ genotype.
- The dominance deviation arises from the property of dominance among alleles at a locus. In absence of dominance, breeding values and genotypic values are equal.

The values for DD in the following table were derived by rearranging the above formula: $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$ to $\delta_{ij} = G_{ij} - \mu - \alpha_i - \alpha_j$.

### Breeding Value / Dominance Deviation

- Genotypic Deviation = genotypic value – population mean
- Breeding Value = sum of effects of individual alleles
- Dominance Deviation = Genotypic deviation – Breeding Value

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic Deviation</th>
<th>Breeding Value</th>
<th>Dominance Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_iA_i$</td>
<td>$2a(a-qd)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_iA_2$</td>
<td>$(q-p)a + 2pqd$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$-2p(a+pd)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E.g. Genotypic Deviation for $A_iA_1 = a \cdot [a(p-q) + 2pq]$.

Then rearranged by $a = \alpha + d(q-p)$. 

85
**Breeding Value / Dominance Deviation**

- Genotypic Deviation = genotypic value – population mean
- Breeding Value = sum of effects of individual alleles
- Dominance Deviation = Genotypic deviation – Breeding Value

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic Value (deviation)</th>
<th>Breeding Value</th>
<th>Dominance Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>2a(a-qd)</td>
<td>2qa</td>
<td>2q²d</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>(q-p)a + 2pqd</td>
<td>(q-p)a</td>
<td>2pqd</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>-2p(a+qd)</td>
<td>-2pa</td>
<td>-2p²d</td>
</tr>
</tbody>
</table>

E.g. Genotypic Deviation for A₁A₁ = \( a - (a(p-q) + 2dpq) \)

Then rearranged by \( a = \alpha + d(q-p) \)
Exercise: Using the previous example: single locus, \( a=1, d=1; p=q=0.5, M= a(p-q) + 2pqd \)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic value</th>
<th>Progeny mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>1</td>
<td>( pa + qd = 1 )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>1</td>
<td>( \frac{1}{2}(p-q)a+d/2 = 0.5 )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>-1</td>
<td>( pd-qa = 0 )</td>
</tr>
</tbody>
</table>

Calculate the breeding value

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Progeny mean</th>
<th>Progeny deviation</th>
<th>Breeding value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Calculate the dominance deviation (=genotypic deviation-BV)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic deviation</th>
<th>Breeding value</th>
<th>Dominance deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph the values to demonstrate BV is a linear regression
**Calculating the progeny mean from parent genotype**

Progeny mean of progenies from parents of different genotype

<table>
<thead>
<tr>
<th>Offspring Genotypes</th>
<th>Parent genotype</th>
<th>A₁A₁</th>
<th>A₁A₂</th>
<th>A₂A₂</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>pa</td>
<td>qd</td>
<td>0</td>
<td>pa+qd</td>
<td></td>
</tr>
<tr>
<td>A₁A₂</td>
<td>(p/2)a</td>
<td>d/2</td>
<td>-(q/2)a</td>
<td>½(p-q)a+d/2</td>
<td></td>
</tr>
<tr>
<td>A₂A₂</td>
<td>0</td>
<td>pd</td>
<td>-qa</td>
<td>pd-qa</td>
<td></td>
</tr>
</tbody>
</table>

Example: single locus, a=1, d=1 (complete dominance for A₁), p=q=0.5

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic value</th>
<th>Progeny mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>1</td>
<td>pa+qd=1</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>1</td>
<td>½(p-q)a+d/2=0.5</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>-1</td>
<td>pd-qa=0</td>
</tr>
</tbody>
</table>

Exercise: Multi-locus- two genotypes, aᵢ=1,dᵢ=1, pᵢ=qᵢ=0.5 for all loci: calculate the genotypic value and progeny mean

i. G₁: A₁A₂, B₁B₂, C₁C₂
ii. G₂: A₁A₁, B₁B₁, C₂C₂

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic value</th>
<th>Progeny mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>G₁</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>G₂</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

The best genotypes do not always make the best parents.

**Average effect of alleles**

Parents pass on alleles – not genotypes – to their offspring. Genotypes are formed anew among the alleles that are passed to the next generation. The values of \(a\) and \(d\) (and \(-a\)) are functions of the genotypes rather than of individual alleles. Therefore they are not useful for expressing the effect of single copy of an \(A₁\) and \(A₂\) allele. The new value associated with the alleles is known as the **average effect of an allele**.

There are several methods by which the average effect could be described; they are equivalent under random mating, but not otherwise.

Fisher’s concept: **The average effect of an allele is the average deviation from the population mean of individuals that received that allele from one parent; the other allele having come from the population at random.** Fisher’s concept of average effect of an allele is the most useful because it is relevant to selection and breeding. Selection aims to change the population mean by favouring one allele over another.
The average effect of $A_1$ allele is denoted as $\alpha_1$ and it is deduced as follows (demonstrate)

<table>
<thead>
<tr>
<th>Type of gamete</th>
<th>Values and frequencies of genotypes produced</th>
<th>Mean value of genotypes produced</th>
<th>Population mean to be deducted</th>
<th>Average effect of an allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1, A_1A_2, A_2A_2$</td>
<td>$a, d, -a$</td>
<td>$pa + qd$</td>
<td>$[a(p-q) + 2dpq]$</td>
<td>$\alpha_1 = q[a + d(q - p)]$</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$p, q$</td>
<td>$pa + qd$</td>
<td>$[a(p-q) + 2dpq]$</td>
<td>$\alpha_1 = q[a + d(q - p)]$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$p, q$</td>
<td>$-qa + pd$</td>
<td>$[a(p-q) + 2dpq]$</td>
<td>$\alpha_2 = -p[a + d(q - p)]$</td>
</tr>
</tbody>
</table>

1. $A_1$ unites at random with another $A_1$ from the population with a frequency of $p$ forms $A_1A_1$
2. $A_1$ unites at random with another $A_2$ from the population with a frequency of $q$ forms $A_1A_2$
3. Multiply frequencies of newly formed genotypes by genotypic values and add for all resulting genotypes = mean value of the genotypes produced
4. Deduct the population mean from the mean value of the genotypes = the deviation from the mean or average effect of an allele
5. The mean average effect of alleles in a population is equal to $pa_1 + qa_2 = 0$
Another concept is the **Average Effect of Allele Substitution**, more convenient when only two alleles are under consideration; defined as the change in the mean of the offspring when the allele is changed to a different allele (e.g. $A_2$ changed to $A_1$).

The average effect of substituting $A_1$ for $A_2$ is denoted by $\alpha$ and it is equal to

$$\alpha = \alpha_1 - \alpha_2 = a + d(q-p)$$

Conversely, the effect of substituting $A_2$ for $A_1$ is equal to $-\alpha$

The average effect of alleles expressed in terms of average effect of allele substitution:

$$\alpha_1 = q\alpha \quad \text{and} \quad \alpha_2 = -p\alpha$$

Note: The average effect of allele substation is a function of:

i) Genotypic values and

ii) Allele frequencies.

**Breeding Value**

We demonstrated that that the average effect of parents’ alleles determine the mean genotypic value of its progeny. The value of an individual, judged by the mean value of its progeny, is called the **breeding value** of the individual. Breeding value can be measured and usually expressed as the deviation from
the population mean. I.e. if an individual is mated to a number of individuals taken at random from the population, its breeding value is **twice** the mean deviation of the progeny from the population. Deviation is doubled because the parent in question provides only half the alleles in the progeny, the other half coming at random from the population.

Defined in the terms of the average effects of alleles, the breeding value of an individual is sum of average effects of alleles that it carries.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Breeding Value</th>
<th>Dominance Deviation</th>
<th>Sum (Genotypic Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1A_1</td>
<td>2α_1=2qα</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_1A_2</td>
<td>α_1 + α_2 = (q-p)α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_2A_2</td>
<td>2α_1= -2pα</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The genotypic value of A_iA_j can be decomposed from \( G_{ij} = \mu + g_{ij} \) to \( G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij} \)

The genotypic value expressed as a deviation from the population mean is \( g_{ij} = \alpha_i + \alpha_j + \delta_{ij} \)

The breeding value associated with A_iA_j is defined as a sum of \( \alpha_i \) and \( \alpha_j \), **sum of effects of individual alleles**.

The dominance deviation arises from the property of dominance among alleles at a locus. In absence of dominance, breeding values and genotypic values are equal.

The concept sum of effects of individual alleles can be easily extended to multiple loci. I.e. the breeding value of genotype A_iA_jB_kB_l is equal to \( \alpha_i + \alpha_j + \alpha_k + \alpha_l \)

**The effect of epistasis on conceptual vs. operational breeding values; discuss in class or homework.**

The term \( \delta_{ij} \) is the dominance deviation associated with the A_iA_j genotype. Dominance deviation, as implied by the term’s name, is a function of dominance at a locus. Dominance affects the genotypic value of heterozygote but not of the homozygotes. Dominance deviation applies both to hetero- and homozygotes. The values for Dominance Deviation in the table below were derived by rearranging the above formula:

\[
G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij} \quad \text{to} \quad \delta_{ij} = G_{ij} - \mu - \alpha_i - \alpha_j
\]

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Breeding Value</th>
<th>Dominance Deviation</th>
<th>Sum (Genotypic Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1A_1</td>
<td>2qα</td>
<td>-2q^2d</td>
<td>2q(α-qd)</td>
</tr>
<tr>
<td>A_1A_2</td>
<td>(q-p)α</td>
<td>2pqd</td>
<td>(q-p)α + 2pqd</td>
</tr>
<tr>
<td>A_2A_2</td>
<td>-2pα</td>
<td>-2p^2d</td>
<td>-2p(α+pd)</td>
</tr>
</tbody>
</table>
Exercise: Using the previous example: single locus, a=1, d=1; p=q=0.5

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Progeny mean</th>
<th>Progeny deviation</th>
<th>Breeding value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1A_1</td>
<td>1</td>
<td>1 - 0.5 = 0.5</td>
<td>1</td>
</tr>
<tr>
<td>A_1A_2</td>
<td>0.5</td>
<td>0.5 - 0.5 = 0</td>
<td>0</td>
</tr>
<tr>
<td>A_2A_2</td>
<td>0</td>
<td>0 - 0.5 = -0.5</td>
<td>-1</td>
</tr>
</tbody>
</table>

Note: Breeding value is twice the deviation of the progeny mean from the population mean.

1. Calculate the breeding value

2. Calculate the dominance deviation=genotypic deviation-BV.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic deviation</th>
<th>Breeding value</th>
<th>Dominance deviation</th>
<th>Genotypic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1A_1</td>
<td>0.5 (=1-0.5)</td>
<td>1</td>
<td>-0.5</td>
<td>1</td>
</tr>
<tr>
<td>A_1A_2</td>
<td>0.5 (=1-0.5)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A_2A_2</td>
<td>-1.5 (= -1-0.5)</td>
<td>-1</td>
<td>-0.5</td>
<td>-1</td>
</tr>
</tbody>
</table>

Note: Genotypic deviation = genotypic value – population mean
Dominance deviation = genotypic deviation – breeding value
Exercise: Graph the values from the table below and draw conclusions.

- Breeding value represents the best linear regression of genotypic deviations on the number of $A_1$ alleles
- Breeding value of a genotype can be predicted from the average of the offspring, no other information is necessary
- Alpha is the best estimate of the effects of substituting an $A_1$ for an $A_2$ allele
- Dominance deviation explain the lack of fit to the linear (additive) model

- $\alpha = a + d(q-p)$
Statistical software workshop

Survey of statistical software use:
- SAS
- JMP
- SPSS
- Agrobase
- Labkey (experiment management)
- Plabstat
- S+
- R
- StatGraph
- In-house
- Others
JMP

- JMP (pronounced jump) is a powerful and interactive data visualization and statistical analysis tool.

- JMP allows will performing analyses and interacting with the data using data tables, graphs, charts, and reports.

- JMP is useful to the researcher who wants to perform a wide range of statistical analyses and modeling.

- JMP is equally useful to the business analyst who wants to quickly uncover trends and patterns in data.

Before you begin using JMP, you should be familiar with these concepts:
- Enter, view, edit, and manipulate data using JMP **data tables**.
- Select a **platform from the Analyze and Graph menus**.

- Platforms use these windows:
  - **Launch windows where you set up and run your analysis**.
  - **Report windows showing the output of your analysis**.

- Report windows normally contain the following items:
  - A graph of some type (such as a scatterplot or a chart).

  - **Specific reports that you can show or hide using the disclosure button**.

  - Platform **options that are located within red triangle menus**.
Differences between Excel and JMP

Formulas
• Excel Formulas are applied to individual cells.
• JMP Formulas are applied only to entire columns.

Columns Names
• Excel Column names are part of the grid. Numbered rows and labeled columns extend past the data.
  — Numeric and character data reside in the same column.

• JMP Column names are not part of the grid. There are no rows and columns beyond the existing data.
  — The grid is only as big as the data. A column is either numeric or character. If a column contains both character and numeric data, the entire column’s data type is character, and the numbers are treated as character data.

Differences between Excel and JMP

Tables and Worksheets
• Excel A single spreadsheet contains several tables, or worksheets.
• JMP JMP does not have the concept of worksheets.
  — Each data table is a separate .jmp file and appears in a separate window.

The Data Grid
• Excel Data can be located anywhere in the data grid.
• JMP Data always begins in row 1 and column 1.

Analysis and Graph Reports
• Excel All data, analyses, and graphs are placed inside the data grid.
• JMP Results appear in a separate window.
JMP data
Each column is formatted in one of the following formats:
- Numeric, continuous
  - data
- Numeric, ordinal
  - Entry numbers, rep number,
  - some discrete data
    - (it will be analyzed by chi square).
- Character, nominal
  - Names, descriptions

JMP
- Data can be completely entered in the JMP table
- Use column panel to manipulate data in entire column
- Use row panel to manipulate data in the entire row
- Or prepare data entirely in excel
- Import data from excel
- Use “Paste with column names”
Choose “Analyze” from the top bar

• Select “Distribution” to analyze single variable
  – E.g. test a mean against a set value

• Select “X by Y” to analyze two variables
  – E.g. regression, correlation

• Select “Fit Model” to analyze multiple variables
  – E.g. ANOVA, t-test for multiple means

JMP help menu
Useful resources – discovering_jmp.pdf (available)

Useful demonstration at:

http://www.youtube.com/watch?v=AKsj0sxCtFA
Quantitative genetic model

1.61.

Advanced models

Phenotypic and Genetic Variances

\[ V_P = V_G + V_E = V_A + V_D + V_I + V_E \]

<table>
<thead>
<tr>
<th>Variance component</th>
<th>Symbol</th>
<th>Variance of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypic</td>
<td>( V_P )</td>
<td>Phenotypic value</td>
</tr>
<tr>
<td>Genetic</td>
<td>( V_G )</td>
<td>Genotypic value</td>
</tr>
<tr>
<td>Additive</td>
<td>( V_A )</td>
<td>Breeding value</td>
</tr>
<tr>
<td>Dominance</td>
<td>( V_D )</td>
<td>Dominance deviation</td>
</tr>
<tr>
<td>Epistatic (interaction)</td>
<td>( V_I )</td>
<td>Epistatic (interaction) effect</td>
</tr>
<tr>
<td>Environment</td>
<td>( V_E )</td>
<td>Environmental deviation</td>
</tr>
<tr>
<td>Genotype X Environment</td>
<td>( V_{GE} )</td>
<td>Genotype x environment interaction</td>
</tr>
<tr>
<td>Error</td>
<td>( V_E )</td>
<td>Experimental error</td>
</tr>
</tbody>
</table>
Additive genetic variance
- Calculated as sum of squared deviations (breeding values)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Breeding Value (Value)²</th>
<th>F*V²</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>p²</td>
<td>2qα</td>
<td></td>
</tr>
<tr>
<td>A₁A₂</td>
<td>2pq</td>
<td>(q-p)α</td>
<td></td>
</tr>
<tr>
<td>A₂A₂</td>
<td>q²</td>
<td>-2pα</td>
<td></td>
</tr>
</tbody>
</table>

VA = σ²ₐ = 4p²q²α² + 2pq(q-p)²α² + 4p²q²α² = 2pqα²(2pq+q²-2pq+p²+2pq) = 2pqα²(p²+2pq+q²)

Exercise: Simplify the formula with assumption:
No dominance (d=0)
Complete dominance (d=a)
p=q=0.5
**Dominance genetic variance**
can be calculated in a similar way.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Dominance Deviation</th>
<th>(DD)²</th>
<th>F*DD²</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>p²</td>
<td>-2q²d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₁A₂</td>
<td>2pq</td>
<td>2pqd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₂A₂</td>
<td>q²</td>
<td>-2p²d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ V_D = \sigma_d^2 = p^2 (-2q^2d)^2 + 2pq (2pqd)^2 + q^2 (-2p^2d)^2 = 4p^2q^2d^2 = (2pqd)^2 \]
Summary of genetic variances

1. $\sigma^2_g = \sigma^2_a + \sigma^2_d$
2. $\sigma^2_a = 2pq[a+d(q-p)]^2$
3. $\alpha = a+d(q-p)$
4. $\sigma^2_d = [2pqd]^2$
5. Relationship between allele frequency and genetic variance
6. Consequences for selection (too early for discussion)
   Understanding the magnitude of $\sigma^2_a$ and $\sigma^2_d$

Covariance between relatives
- The covariance between relatives measures the degree of genetic resemblance between related individuals in a population.
- By definition the covariance between unrelated individuals is zero.
- The covariance between relatives underlies the effectiveness of selection in breeding programs.
- The progress from selection is directly proportional to the degree of resemblance between the selected individuals and their progenies.
- The covariance between relatives can be used for estimating the genetic variances.

Statistically, covariance is a measure of association between two variables (values without subscript are means):

$$\text{Cov}_{XY} = \Sigma_{i=1}^{n}[(X_i - \bar{X})(Y_i - \bar{Y})]/(N-1)$$
Covariance of Offspring and Parent

X: Parent genotypic value as deviation from the population mean
Y: Offspring mean genotypic value as deviation from the population mean

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Parent Genotypic value</th>
<th>Offspring Mean genotypic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>p²</td>
<td>2q(α-qd)</td>
<td></td>
</tr>
<tr>
<td>A₁A₂</td>
<td>2pq</td>
<td>(q-p)α + 2pqd</td>
<td></td>
</tr>
<tr>
<td>A₂A₂</td>
<td>q²</td>
<td>-2p(α+pd)</td>
<td></td>
</tr>
</tbody>
</table>

Σᵢ₌₁ⁿ XᵢYᵢ → Cov_{OP} = [p²q(α-qd)qα] + [2pq((q-p)α+2pqd)(q-p)α/2] + [q²(-2p(α+pd)(-pα))= 
pqa²(p²+2pq+q²)+2p²q²αd(-q+q-p+p) = 
= pqα² = ½σₐ²

The value applies to both offspring to one- and mid-parent
Half sibs

• One parent in common
• The other parent is at random from the population

• By definition – the mean genotypic value of the group of half sibs is half the breeding value of the common parent

• The covariance is the variance of the true means of the half-sib groups, therefore the variance of the half the breeding values of the common parent, which equals $\frac{1}{4} V_A$

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Deviation – Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$p^2$</td>
<td>$q\alpha$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$2pq$</td>
<td>$(q-p)\alpha/2$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$q^2$</td>
<td>$-p\alpha$</td>
</tr>
</tbody>
</table>

\[
\text{Var}_{HS} = p^2q^2\alpha^2 + 2pq \frac{1}{2} (q-p)^2\alpha^2 - q^2p^2\alpha^2 = pq\alpha^2 \left[ \frac{1}{2} (q-p)^2 \right] = \frac{1}{2} pq\alpha^2 = \frac{1}{4} \sigma_a^2.
\]

From previous $pq\alpha^2 = \frac{1}{2} \sigma_a^2$, therefore $pq\alpha^2 = \frac{1}{4} \sigma_a^2$.  

104
Full sibs

- Both parents are common

- The covariance is the variance of the means of full sib families.
  - Calculation in table 9.2 (Falconer and Mackay, page 149)

- $\text{Cov}_{(FS)} = \frac{1}{2} V_A + \frac{1}{4} V_D$

<table>
<thead>
<tr>
<th></th>
<th>Coefficient for $V_A$</th>
<th>Coefficient for $V_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual with itself</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Parent-offspring</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>Half-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred common parent</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
</tr>
<tr>
<td>Full-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred parents</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>
**Covariance between relatives common in plant breeding**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient for $V_a$</th>
<th>Coefficient for $V_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual with itself</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Parent-offspring</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>Half-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred common parent</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
</tr>
<tr>
<td>Arbitrary $F$</td>
<td>$(1+F)/4$</td>
<td></td>
</tr>
<tr>
<td>Full-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred parents</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>Arbitrary $F_a$ and $F_b$</td>
<td>$(2+F_a+F_b)/4$</td>
<td>$[(1+F_a)(1+F_b)]/4$</td>
</tr>
</tbody>
</table>

**Summary of Quantitative Genetic Model**

- **Population mean:**
  - Values, allele and genotype frequencies
  - Phenotypic and genotypic values, env. deviation

- **Genotype - Progeny Mean**
  - Genotypic value – not a good predictor of Progeny Mean

- **Average effect of alleles**
  - the average deviation from the population mean of individuals that received that allele from one parent

- **With no dominance**
  - Breeding Value = Genotypic Value

- **Breeding Value**
  - Sum of average effects of alleles that it carries

- **Genotypic Value – Breeding Value = Dominance Deviation**

- **Variances (concept):**
  - Additive (BV), dominance, interaction, environmental
  - Deviation squared x frequencies

- **Variances (how to measure):**
  - Resemblance between relatives, explore covariance-variance relationships
  - Parent-Offspring, Half-Sibs, Full-Sibs
Additional Reading

Page 16 in notes.

Falconer & Mackay, 1996
Chapters 1-5 – Population genetics
Chapters 6-9 – Quantitative genetics

Recommended text books:
http://stemmapress.com
Phenotypic and Genetic Variances

As discussed previously – phenotypic values are due to genetic and environmental effects. The genetics of a metric character is based on the study of its variation. This is done by partitioning the phenotypic variance into different causal components. The relative magnitude of these components is determined by the genetic properties of a population, and in particular the resemblance between relatives. The components of the variance are listed in the table below. The phenotypic variance is a sum of the components: $V_P = V_G + V_E = V_A + V_D + V_I + V_E$

<table>
<thead>
<tr>
<th>Variance component</th>
<th>Symbol</th>
<th>Variance of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypic</td>
<td>$V_P$</td>
<td>Phenotypic value</td>
</tr>
<tr>
<td>Genetic</td>
<td>$V_G$</td>
<td>Genotypic value</td>
</tr>
<tr>
<td>Additive</td>
<td>$V_A$</td>
<td>Breeding value</td>
</tr>
<tr>
<td>Dominance</td>
<td>$V_D$</td>
<td>Dominance deviation</td>
</tr>
<tr>
<td>Epistatic (interaction)</td>
<td>$V_I$</td>
<td>Epistatic (interaction) effect</td>
</tr>
<tr>
<td>Environment</td>
<td>$V_E$</td>
<td>Environmental deviation</td>
</tr>
<tr>
<td>Genotype X Environment</td>
<td>$V_{GE}$</td>
<td>Genotype x environment interaction</td>
</tr>
<tr>
<td>Error</td>
<td>$V_{E}$</td>
<td>Experimental error</td>
</tr>
</tbody>
</table>

The partition of the variance allows estimation of the relative importance of the various components (determinants) on the phenotype.

Most importantly for breeders, it allows the estimation of heritability.

Heritability = ratio of the genetic component to the total phenotypic variance, i.e. $V_G/V_P$ it is called **heritability in broad sense or coefficient of genetic determination**

$$h^2 = V_A/V_P = \textit{heritability in a narrow sense}$$

Digression – Statistics: Variance

$Y_i$ = individual trait value, $Y$ = trait mean in a population, $N$ is the number if individuals in the population

$(Y_i - \bar{Y}) = \text{individual deviation from the mean}$

$(Y_i - \bar{Y})^2 = \text{individual squared deviation from the mean}$

$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \text{sum of squared deviation from the mean}$

$\sum_{i=1}^{n} (Y_i - \bar{Y})^2/N = \text{mean of squared deviation from the mean = the Variance} = \sigma_Y^2$

Variance based on frequencies: $\sigma_Y^2 = \sum_{i=1}^{n} f_i (Y_i - Y)^2$, or $\sum_{i=1}^{n} f_i (Y_i)^2 = F*V^2$ if values are eYpressed as deviations
The additive genetic variance can be calculated from breeding value as deviation from the population mean. Consider a single locus with alleles A₁ and A₂ and frequencies p and q. Recall the breeding values for different genotypes. The variance of these breeding values is the sum of the products of the genotype frequencies and the squared breeding values.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Breeding Value</th>
<th>(Value)²</th>
<th>F*V²</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>p²</td>
<td>2qα</td>
<td>4q²α²</td>
<td>4p²q²α²</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>2pq</td>
<td>(q-p)α</td>
<td>(q-p)²α²</td>
<td>2pq(q-p)²α²</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>q²</td>
<td>-2pa</td>
<td>4p²α²</td>
<td>4p²q²α²</td>
</tr>
</tbody>
</table>

\[ V_A = \sigma_a^2 = 4p^2q^2\alpha^2 + 2pq(q-p)\alpha^2 + 4p^2q^2\alpha^2 = 2pq\alpha^2(2pq+q^2-2pq+p^2+2pq) = 2pq\alpha^2(p^2+2pq+q^2) = 2pq\alpha^2 = 2pq[a + d(q-p)]^2 \]

Exercise: Simplify the formula with assumption:

1. No dominance (d=0)
2. Complete dominance (d=a)
3. p=q=0.5

Dominance genetic variance can be calculated in a similar way.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Dominance Deviation</th>
<th>(DD)²</th>
<th>F*DD²</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁A₁</td>
<td>p²</td>
<td>-2q²d</td>
<td>(-2q²d)²</td>
<td>p² (-2q²d)²</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>2pq</td>
<td>2pqd</td>
<td>(2pqd)²</td>
<td>2pq (2pqd)²</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>q²</td>
<td>-2p²d</td>
<td>(-2p²d)²</td>
<td>q² (-2p²d)²</td>
</tr>
</tbody>
</table>

\[ V_d = \sigma_d^2 = p^2 (-2q²d)² + 2pq (2pqd)² + q² (-2p²d)² = 4p^2q^2d^2 = (2pqd)² \]

Summary of genetic variances

1. \( \sigma_g^2 = \sigma_a^2 + \sigma_d^2 \)
2. \( \sigma_a^2 = 2pq[a+d(q-p)]^2 \)
3. \( \alpha = a+d(q-p) \)
4. \( \sigma_d^2 = [2pqd]^2 \)
5. Relationship between allele frequency and genetic variance
6. Consequences for selection

Exercise: Discuss the graphs showing the magnitude of genetic component in relation to allele frequencies (Falconer & Mackay, page 128)
Covariance between relatives

Resemblance between relatives is one of the earliest observations. Close relatives such as parent and offspring have a greater resemblance than more distant relatives, e.g. uncle and niece.

Falconer makes a differentiation between partitioning of the phenotypic variance into *casual components of variance* ($V$) and partitioning of the phenotypic variance, into components corresponding to the grouping of individuals into families, *observational components of phenotypic variance* ($\sigma^2$).

The covariance between relatives measures the degree of genetic resemblance between related individuals in a population. By definition the covariance between unrelated individuals is zero.

Non-genetic factors could also contribute to resemblance between relatives, but non-genetic factors among relatives are assumed to be uncorrelated. This assumption is met by through randomization procedures in the experimental designs. The

The covariance between relatives underlies the effectiveness of selection in breeding programs. The progress from selection is directly proportional to the degree of resemblance between the selected individuals and their progenies. The covariance between relatives can be used for estimating the genetic variances.

Covariance of Offspring and Parent

Covariance is a measure of association between two variables (a pair of traits, values of pairs of sibs, values for parent and offspring,...)

$$\text{Cov}_{XY} = \Sigma^{n}_{i=1}[(X_i-\bar{X})(Y_i-\bar{Y})]/(N-1) = \Sigma^{n}_{i=1}fX_iY_i - \bar{X}\bar{Y}$$ (or $\Sigma^{n}_{i=1}fX_iY_i$ if values are deviations)

### Offspring and parent values

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Deviation - Parent</th>
<th>Deviation – Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td>$p^2$</td>
<td>$2q(\alpha - qd)$</td>
<td>$qa$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>$2pq$</td>
<td>$(q-p)\alpha + 2pqd$</td>
<td>$(q-p)\alpha/2$</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>$q^2$</td>
<td>$-2p(\alpha + pd)$</td>
<td>$-p\alpha$</td>
</tr>
</tbody>
</table>

Offspring deviations are determined as deviations from offspring means, which without selection, equal population means.

Use $\Sigma^{n}_{i=1}fX_iY_i \rightarrow \text{Cov}_{OP} = [p^22q(\alpha - qd)qa] + [2pq((q-p)\alpha + 2pqd)(q-p)\alpha/2] + [q^2(-2p(\alpha + pd)(-p\alpha)] = pq\alpha^2(p^2+2pq+q^2)+2p^2q^2\alpha(-q+q-p+p) = pq\alpha^2 = \frac{1}{2}\alpha^2$

The same covariance is calculated for both offspring to one parent and offspring to mid-parent

Half Sibs
One parent in common, the other parent is at random from the population
By definition – the mean genotypic value of the group of half sibs is half the breeding value of the common parent
The covariance is the variance of the true means of the half-sib groups, therefore the variance of the half the breeding values of the common parent, which equals \( \frac{1}{4} \sigma_A^2 \).

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Deviation – Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1A_1)</td>
<td>(p^2)</td>
<td>(q\alpha)</td>
</tr>
<tr>
<td>(A_1A_2)</td>
<td>(2pq)</td>
<td>((q-p)\alpha/2)</td>
</tr>
<tr>
<td>(A_2A_2)</td>
<td>(q^2)</td>
<td>(-p\alpha)</td>
</tr>
</tbody>
</table>

\[
Var_{HS} = p^2q^2\alpha^2 + 2pq \frac{1}{2} (q-p)^2\alpha^2 - q^2p^2\alpha^2 = pqa^2[(p+q)\frac{1}{2}(q-p)^2] = \frac{1}{2} pqa^2 = \frac{1}{4} \sigma_a^2
\]

From previous \( pqa^2 = \frac{1}{2} \sigma_a^2 \); therefore \( \frac{1}{2} pqa^2 = \frac{1}{4} \sigma_a^2 \)

\[
Cov_{HS} = Var_{HS} = Var_{1/2A} = \frac{1}{4} Var_A
\]

**Full Sibs**

\[
Cov_{FS} = \text{Var}_{FS \text{ (family means)}} = \text{Var}_{(A+D)(A+D)} = \text{Var}_{(AA)} + \text{Var}_{(DD)} = \frac{1}{2} Var_A + \frac{1}{4} Var_D
\]

**Covariance between relatives common in plant breeding**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient for (V_A)</th>
<th>Coefficient for (V_D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual with itself</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Parent-offspring</td>
<td>(\frac{1}{2})</td>
<td>0</td>
</tr>
<tr>
<td>Half-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred common parent</td>
<td>(\frac{1}{4})</td>
<td>0</td>
</tr>
<tr>
<td>Arbitrary (F)</td>
<td>((1+F)/4)</td>
<td></td>
</tr>
<tr>
<td>Full-sibs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-inbred parents</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{4})</td>
</tr>
<tr>
<td>Arbitrary (F_A) and (F_B)</td>
<td>((2+F_A+F_B)/4)</td>
<td>([(1+F_A)(1+F_B)]/4)</td>
</tr>
</tbody>
</table>

**Multilocus extension.**

1. If the loci act independently, the variances for the trait are the sum of the single-locus contributions.
2. Deviations from the assumption of independence:

   a. Epistasis
      i. Genotypic value = Breeding value + Dominance deviation + Interaction = \(A+D+I\)
      ii. \(\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_I^2\)
      iii. Summary of higher order terms for locus pairs

<table>
<thead>
<tr>
<th>Genetic Contribution</th>
<th>Name</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allelic Type</td>
<td>Genetic Effect</td>
<td>Symbol</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Single alleles (one locus)</td>
<td>Additive</td>
<td>A</td>
</tr>
<tr>
<td>Pairs of alleles (one locus)</td>
<td>Dominance</td>
<td>D</td>
</tr>
<tr>
<td>Non-allelic pairs (separate loci)</td>
<td>Additive x Additive</td>
<td>AA</td>
</tr>
<tr>
<td>Single allele with allelic pair (separate loci)</td>
<td>Additive x Dominance</td>
<td>AD</td>
</tr>
<tr>
<td>Two allelic pairs (separate loci)</td>
<td>Dominance x Dominance</td>
<td>DD</td>
</tr>
</tbody>
</table>

**Why all this? Do breeders need to estimate genetic variances?**

- Usefulness criterion requires information on both the mean and genetic variance in a population.
- Estimation of means is straightforward, therefore choosing population with a high mean is routine
- Estimation of variances is more difficult for two reasons
  - second order statistics = squared effect = higher sampling error than for means
  - require mating design which are not routinely used in inbred or hybrid development program

Mating design – a topic for week 2

**Usefulness criterion** combines information on the mean performance and genetic variance in a population. Usefulness criterion ($U_p$) is equal to expected genotypic mean of selected proportion of individuals in a population.

\[ U_p = \mu_{RI} + k_p \left( \frac{2V_A}{\sigma V_P} \right) \]

\[ U_p = \mu_T + k_p \left( \frac{2V_A}{\sigma V_P} \right) \]

$\mu_{RI}$ = mean of recombinant inbreds; $k_p$ = standardized selection differential; $2V_A$ = genetic variance among recombinant inbreds; $V_P$ = phenotypic variance among recombinant inbreds.

For recombinant inbred lines (RIL) $U_{p(RI)} = \mu_{RI} + k_p \left( \frac{2V_A}{\sigma V_P} \right)$

For testcrosses $U_{p(T)} = \mu_T + k_p \left( \frac{2V_A}{\sigma V_P} \right)$
Summary of Quantitative Genetic Model

Population mean:
Values, allele and genotype frequencies
Phenotypic and genotypic values, env. deviation

Genotype - Progeny Mean
Genotypic value – not a good pre

Average effect of alleles
the average deviation from the population mean of
individuals that received that allele from one parent

Breeding Value
Sum of average effects of alleles t

With no dominance
Breeding Value = Genotypic Value

Genotypic Value – Breeding Value

Variances (concept):
Additive (BV), dominance, interaction, environmental
Deviation squared x frequencies

Variances (how to measure): Re
relatives, explore covariance-var
Parent-Offspring. Half-Sibs. Full
Summary of Quantitative Genetic Model

0. Values, \( a, d, -a \)

1. Population mean = \( \mu = \Sigma F^*V = p^2a+2pqd-q^2a = a(p-q) + 2pqd \)

2. Average Effect of an Allele

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Mean Value of genotype</th>
<th>Population Mean</th>
<th>Average Effect of an Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( p )</td>
<td>( pa + qd )</td>
<td>( [a(p-q) + 2pqd] )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( q )</td>
<td>( -qa + pd )</td>
<td>( [a(p-q) + 2pqd] )</td>
</tr>
</tbody>
</table>

3. The average effect of substituting \( A_1 \) for \( A_2 \) is denoted by \( \alpha \) and it is equal to:

\[
\alpha = \alpha_1 - \alpha_2 = a + d(q-p)
\]

The average effect of alleles expressed in terms of average effect of allele substitution:

\[
\alpha_1 = q\alpha; \quad \alpha_2 = -p\alpha
\]

4. Breeding Value

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Sum (Genotypic Deviation)</th>
<th>Breeding Value</th>
<th>Dominance Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( 2q(a-qd) )</td>
<td>( 2qa ) or ( 2\alpha_1 )</td>
<td>(-2q^2d )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( (q-p)a + 2pqd )</td>
<td>( (q-p)a ) or ( \alpha_1 + \alpha_2 )</td>
<td>( 2pqd )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>(-2p(a+pd) )</td>
<td>(-2p\alpha ) or ( 2\alpha_2 )</td>
<td>(-2p^2d )</td>
</tr>
</tbody>
</table>

Genotypic Deviation = gen value – pop mean = \( a - \left[ a(p-q) + 2pqd \right] \)

5. The additive genetic variance

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Breeding Value</th>
<th>((\text{Value})^2)</th>
<th>(F^*V^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( p^2 )</td>
<td>( 2qa )</td>
<td>( 4q^2\alpha^2 )</td>
<td>( 4p^2q^2\alpha^2 )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( 2pq )</td>
<td>( (q-p)a )</td>
<td>( (q-p)^2\alpha^2 )</td>
<td>( 2pq(q-p)^2\alpha^2 )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>( q^2 )</td>
<td>(-2p\alpha )</td>
<td>( 4p^2\alpha^2 )</td>
<td>( 4p^2q^2\alpha^2 )</td>
</tr>
</tbody>
</table>

\( V_A = \sigma^2a = \sum F^*V^2 = \sum 4p^2q^2\alpha^2 - 2pq[a + d(q-p)]^2 \)

6. The dominant genetic variance

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Dominance Deviation</th>
<th>((DD)^2)</th>
<th>(F^*DD^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( p^2 )</td>
<td>(-2q^2d )</td>
<td>((-2q^2d)^2 )</td>
<td>( p^2(-2q^2d)^2 )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( 2pq )</td>
<td>( 2pqd )</td>
<td>((2pqd)^2 )</td>
<td>( 2pq(2pqd)^2 )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>( q^2 )</td>
<td>(-2p^2d )</td>
<td>((-2p^2d)^2 )</td>
<td>( q^2(-2p^2d)^2 )</td>
</tr>
</tbody>
</table>

\( V_D = \sigma^2d = \sum F^*DD^2 = \sum 4p^2q^2d^2 - (2pqd)^2 \)

7. Offspring and parent values

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Deviation - Parent</th>
<th>Deviation – Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 )</td>
<td>( p^2 )</td>
<td>( 2q(a-qd) )</td>
<td>( qa )</td>
</tr>
<tr>
<td>( A_1A_2 )</td>
<td>( 2pq )</td>
<td>( (q-p)a + 2pqd )</td>
<td>( (q-p)a/2 )</td>
</tr>
<tr>
<td>( A_2A_2 )</td>
<td>( q^2 )</td>
<td>(-2p(a+pd) )</td>
<td>( -pa )</td>
</tr>
</tbody>
</table>

Offspring deviations are determined as deviations from offspring means, which without selection, equal population means.

Use \( \Sigma_{i=1}^n f_i Y_i \rightarrow \text{Cov}_{OP} = [p^2 2q(a-qd)qa] + [2pq((q-p)a+2pqd)(q-p)a/2] + [q^2(-2p(a+pd)(-pa)] = pq\alpha^2(p^2+2pq+q^2)+2p^2q^2d(-q+q+p+p) = pq\alpha^2 = \frac{1}{2}\sigma^2_a \)
PERSPECTIVES

Survey of U.S. Land-Grant Universities for Training of Plant Breeding Students

Nihat Guner and Todd C. Wehner

ABSTRACT

A survey was conducted to identify land-grant universities in the USA having plant breeding programs, and to determine the number of domestic and international plant breeding students graduating at the M.S. and Ph.D. levels from those programs in 1995 to 2000. A total of 71 U.S. land-grant universities were identified. There were 409 (53%) Ph.D. and 361 (47%) M.S. degrees awarded. Of the total, 362 (47%) graduates were domestic and 408 (53%) were international. There was no major change in the total number of plant breeding graduates during the 6-yr period. The largest numbers of plant breeding students were trained in agronomy (crop science) departments, followed by plant breeding departments or groups, horticulture departments, plant science departments, and combined agronomy–horticulture departments. Universities with an average of seven or more graduates per year were University of Wisconsin-Madison, North Carolina State University, University of Nebraska-Lincoln, Cornell University, University of Minnesota-St. Paul, Iowa State University, and Texas A&M University. The downward trend noted in previous surveys has continued to the point where there are only a few universities with large plant breeding programs remaining in each region of the country. If the USA is going to continue its public effort in plant breeding research and graduate student training, sufficient federal and state funding will have to be provided to support at least the current regional centers.

FOR THE PAST SEVERAL YEARS, there have been concerns in the public and private plant breeding sectors about the future availability of plant breeding graduates to work at seed companies and to fill other plant breeding positions (Collins and Phillips, 1991). Traditional plant breeding programs in U.S. land-grant universities have been diminishing as financial and human resources are reduced. Funding for plant breeding research and development today is in short supply in land-grant universities. From 1980 to the mid-1990s, 30 plant breeding positions (6% of the total) associated with graduate programs were eliminated. This downward trend has continued (Frey, 2000). Fewer land-grant universities are now involved in training students, and the remaining breeders are working with dwindling government support. The number of public plant breeding programs has declined during the previous few decades due to a lack of funds for agricultural research and to a redirection of funding to research in molecular genetics (Frey, 1996). Fewer plant breeders, working in fewer universities, are now providing the education and training for students interested in plant breeding. With a reduction in the number of training programs comes an increased need for planning to increase the number of plant breeding graduates. Planning is also needed to provide private industry with information on where its future workforce will come from.

Public sector agricultural research in general, and public plant breeding research in particular, is in trouble in both industrialized and developing countries. Budgets have leveled off, or are declining in the public sector. During the last 30 yr or more, private sector agricultural research investment has grown dramatically, and an increasing proportion of this investment has been directed to plant breeding (Heisey et al., 2001). Heisey et al. (2001) indicated that the main reasons for declining financial support in public plant breeding are scientific advance and the cost of research innovations.

Statistical data on federal funding for agricultural research and development show that government support in the USA has declined dramatically during the last half century. In 1940, nearly 40% of federal funding for research and development was allocated to agricultural research and development (Mowery and Rosenberg, 1989). By 1991, the USDA expenditure for research and development was only 2% of total federal research and development spending, and only 4% of federal research support at colleges and universities was devoted to agricultural research and development (Fugile et al., 1996). Federal funding for agricultural research and development increased at an annual rate of 2.8% until 1980, after which it has been nearly level (Frey, 2000). Several surveys have been conducted during the past 17 yr to determine the number of plant breeding graduates from universities in the USA. Brooks and Vest (1985) conducted a survey of public programs on genetics and breeding of horticultural crops. The survey of 98 institutions projected a 13% decline in the number of horticultural plant breeders in the public sector from 1983 to 1990.

James (1990) reported that the U.S. public sector supported 417 plant breeders in 1989. That study reported 144 plant breeders in horticultural crop breeding, which was 19.6% less than the 179 horticultural crop breeders reported by Brooks and Vest (1985) in 1983. James (1990) pointed out that the number of plant breeding programs, as well as the number of plant breeders, had been declining across time. He suggested that programs on the improvement of minor crops would be most severely affected.

Collins and Phillips (1991) surveyed all public land-
grant universities and 1890 universities on graduate training in plant breeding in the USA. They reported the number of plant breeders working on all crops was 477 in 1980, declining to 459 in 1989. This result showed that the loss of plant breeders was 1.8 per year during the 1980s. During the 10-yr period, the number of plant breeding programs decreased 9.3%, and the number of plant breeders decreased 3.9%. Collins and Phillips (1991) also indicated that changes in most states were relatively small, with California reporting the largest reduction in the number of plant breeding programs. However, a few states showed increases. During the 1980s, land-grant universities in Wisconsin, North Carolina, Texas, and Florida had the largest plant breeding programs.

The decline in public plant breeding programs has occurred in part because of the increase in private plant breeding programs. Kalton and Richardson (1983) surveyed private companies having plant breeding programs and reported that there were 1191 plant breeders in 1981. Kalton et al. (1989) conducted a second survey of private companies in 1988 and reported 1568 plant breeders, a 32% increase.

Frey (1996) conducted a survey on plant breeding research and development in the public sector as part of the U.S. National Plant Breeding Study and reported that the public sector decreased 2.5 scientist-years per year from 1990 to 1994. During the same period, private industry was found to have an annual net growth of 32 scientist-years with a gain of 2.4% per year. In a second study, Frey (1997) reported that the private breeding sector spent $338 million annually on research and development, 61% of the U.S. annual expenditure on research and development of $551 million, in 1994.

Collins and Phillips (1991) and Frey (1996) were concerned that the number of public plant breeders available to educate future plant breeders was on a slow and steady decline in the USA. From 1980 to 1994, there was a loss of 30 (full-time equivalent) public plant breeders (Frey, 1996). While oversupply of students can create problems for graduating students looking for appropriate jobs, the decline in public plant breeders along with the increase in private plant breeders appears to represent the opposite problem. To plan properly, universities should count the number of plant breeding students being trained, as well as plant breeders being hired by employers such as universities, USDA laboratories, seed companies, and food processors. Undergraduate students interested in graduate training in plant breeding need to know which universities have programs, and which programs are large enough to permit the teaching of courses in the various sub-areas of plant breeding. Small university programs are useful for training in niche areas.

Numerous surveys conducted between 1983 and 1997 indicated that there was a declining number of public plant breeders, and an increasing number of private plant breeders. Our survey was initiated to assess the current situation of plant breeding training in the public sector in the USA. The main objective of this survey was to identify land-grant universities in the USA that have plant breeding programs, and to determine the numbers of plant breeding students at the M.S. and Ph.D. levels graduating from 1995 to 2000. We were also interested to know whether the students were in agronomy or horticulture, and whether they were domestic or international.

**MATERIALS AND METHODS**

In 2001, survey questionnaires were sent to all land-grant universities in the USA that were thought to have plant breeding programs. A total of 71 land-grant universities were identified in the USA. Of those, 51 had faculty and/or courses that could be used to offer plant breeding training leading to the M.S. or Ph.D. degrees. In the survey questionnaire, information was requested on the numbers of students trained from 1995 to 2000 in M.S. and Ph.D. levels for domestic and international plant breeding graduates. No sampling techniques were used since we counted every land-grant university. However, universities were asked to make a judgment when sending us the count of plant breeding students. We asked that they send us the count excluding graduate students working mostly in molecular genetics. It should be noted that the number of plant breeding degrees awarded is probably larger than the number of students graduating, since some students continue for a Ph.D. after receiving their M.S. degree.

The survey questionnaires were sent directly to departments in which plant breeding graduates could be trained. In land-grant universities, student counts of plant breeding graduates were requested from departments of horticulture or agronomy (crop science), as well as from combined agronomy and horticulture departments or general plant science departments. There were only two universities having either an official plant breeding department (Cornell University) or interdepartmental plant breeding program (University of Wisconsin-Madison). However, most large departments of agronomy or horticulture had a plant breeding group and/or a specialized plant breeding curriculum. Universities were assigned to a region of the USA based on the Cooperative State Research, Education, and Extension Service classification.

Information about the universities was obtained by university catalogs available in the library, as well as from the world wide web. We identified those universities having departments of horticulture, agronomy (crop science), plant science, combined agronomy and horticulture, or plant breeding. We then identified departments that offered courses or other training in plant breeding. Surveys were sent first by electronic mail (email) or telephone-transmitted facsimile (fax). The survey was resent to universities not responding to the first request. Telephone calls were used as a follow up for those not responding after three requests, and to clarify answers to the survey questions. Data obtained from the survey were summarized as means and frequencies.

**RESULTS AND DISCUSSION**

Responses were received from 78 departments offering plant breeding degrees representing 47 land-grant universities in 47 states. Seven states reported no degree programs in plant breeding and four states did not respond to survey. This resulted in a 95% return rate.

**Number of Plant Breeding Graduates**

During the survey years (1995 to 2000), 770 graduate degrees in plant breeding were awarded from 82 depart-
mments located at 47 land-grant universities (Table 1). There were 409 (53%) Ph.D. and 361 (47%) M.S. degrees awarded. Of the total, 362 (47%) graduates were domestic and 408 (53%) were international. Many of the universities reported having undergraduate programs, or at least courses, in plant breeding. However, most students entering graduate programs in plant breeding obtained their bachelor’s degree in a biological science department including agronomy (or crop science), biology, botany, horticulture, or forestry.

Changes in the number of plant breeding degrees awarded during the 6-yr survey period were relatively small, with 1997 having the highest number of graduates and 1998 having the lowest (Fig. 1). The total number of graduates in 1995 was similar to that in 2000. Overall, 1996 and 1997 had the most graduates, but otherwise the numbers were relatively constant. The same general trend can be seen in the numbers of M.S. and Ph.D. degrees awarded. The numbers of M.S. and Ph.D. graduates in 1995 and 2000 were similar.

For student origin, there were only small changes in the number of domestic and international plant breeding graduates across years. In 1995, the numbers of domestic and international graduates were similar. However, the total number of international graduates increased in 1996 and 1997. The trend then reversed, and in 2000, there were slightly more domestic graduates than international graduates (Fig. 1).

For degree, the trend was different than for total. The number of international Ph.D. graduates was significantly higher than the number of domestic Ph.D. graduates, as was the case in each year (Fig. 2). The number of international Ph.D. graduates increased greatly in 1996 and 1997. In general, the total number of international plant breeding Ph.D. graduates increased across years, while the total number of domestic Ph.D. graduates did not change much. There was a decrease in domestic Ph.D. graduates in 1999, but the number bounced back in 2000.

**Major Centers for Plant Breeding Training**

On the basis of the number of plant breeding students trained (Fig. 3), the top seven universities involved in plant breeding training were the University of Wisconsin-Madison, North Carolina State University, University of Nebraska-Lincoln, Cornell University, University of Minnesota-St. Paul, Iowa State University, and Texas A&M University. The geographical distribution of the top universities shows that the Midwest is a major region for the training of plant breeders in the USA (Fig. 4). Thus, each region of the USA had at least one major university offering plant breeding: Cornell University in the Northeast, North Carolina State University in the Southeast, University of Wisconsin-Madison (along with University of Minnesota-St. Paul and Iowa State University) in the Midwest, University of Nebraska-Lincoln (along with North Dakota State University) in the Great Plains, Texas A&M University in the South-central/Southwest, and Oregon State University in the West (Fig. 4). The major universities involved in plant breeding training have large programs that enable students to get a wide range of courses and practical training in the various areas of plant breeding, attracting a large number of students from the USA and abroad.

**Departments Offering Plant Breeding Training**

A total of 82 departments located in 51 land-grant universities offer plant breeding degrees and were surveyed in this study. Of all the departments surveyed, most were separate departments of agronomy (or crop science) and horticulture, each representing 34% of all departments (Fig. 5). Fewer departments were plant science (21%) or combined agronomy and horticulture (5%). Apparently, plant breeding programs operate

<p>| Table 1. Numbers of plant breeding students trained at land-grant universities in the USA from 1995 to 2000.† |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Land-grant university</th>
<th>Total</th>
<th>Domestic</th>
<th>International</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>770</td>
<td>153</td>
<td>207</td>
</tr>
<tr>
<td>Univ. of Wisconsin-Madison</td>
<td>72</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>North Carolina State Univ.</td>
<td>64</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>Univ. of Nebraska-Lincoln</td>
<td>61</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Cornell Univ.</td>
<td>54</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Univ. of Minnesota</td>
<td>48</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Iowa State Univ.</td>
<td>44</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Texas A&amp;M Univ.</td>
<td>41</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Michigan State Univ.</td>
<td>36</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>North Dakota State Univ.</td>
<td>33</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Oregon State Univ.</td>
<td>28</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>New Mexico State Univ.</td>
<td>25</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>South Dakota State Univ.</td>
<td>23</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Kansas State Univ.</td>
<td>20</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mississippi State Univ.</td>
<td>20</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Univ. of Puerto Rico</td>
<td>17</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Washington State Univ.</td>
<td>17</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Univ. of Florida</td>
<td>15</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Auburn Univ.</td>
<td>15</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Univ. of Idaho</td>
<td>15</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Univ. of Illinois-Urbana</td>
<td>14</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Purdue Univ.</td>
<td>14</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Univ. of Missouri</td>
<td>13</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Virginia Polytechnic Inst. &amp; State Univ.</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Louisiana State Univ.</td>
<td>8</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Univ. of Hawaii</td>
<td>8</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Arkansas</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Univ. of Maryland</td>
<td>6</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ohio State Univ.</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Colorado State Univ.</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pennsylvania State Univ.</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Univ. of California-Davis</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Univ. of California-Riverside</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Clemson Univ.</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Georgia</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Univ. of Kentucky</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Connecticut</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Delaware</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of New Hampshire</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Univ. of Wyoming</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rutgers Univ.</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Alabama A&amp;M Univ.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oklahoma State Univ.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tuskegee Univ.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Univ. of Alaska-Fairbanks</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Guam</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Univ. of Maine</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Rhode Island</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Tennessee</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Univ. of Vermont</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Virginia State Univ.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>West Virginia Univ.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

† Counts do not include students with primary training in molecular genetics.
‡ Not responding to survey, but probably zero.
mainly in departments large enough to support a diversity of disciplines, mostly excluding the combined or general plant science departments having fewer agronomy or horticulture faculty.

The largest number of plant breeding students were trained in agronomy (crop science) departments, followed by plant breeding departments or groups, horticulture departments, plant science departments, and combined agronomy and horticulture departments (Fig. 6). Although plant breeding departments or groups were only a small percentage of all departments, they were the second most important source of plant breeding student training.

**Trends in Plant Breeding Graduates**

Survey results showed that there were no major changes in the total number of plant breeding graduates during the past 6 yr in the USA, with only 2% more in 2000 than in 1995. The numbers fluctuated across years,
with an increase in the numbers of plant breeding graduates during 1996 and 1997. Some changes have occurred year-to-year in terms of the total numbers of domestic and international plant breeding graduates in the levels of M.S. and Ph.D. The number of domestic Ph.D. graduates remained steady, while the number of international Ph.D. graduates increased slightly. The number of domestic M.S. graduates had a large increase, while the number of international M.S. graduates declined.

A plant breeding training survey of public institutions was conducted by Collins and Phillips (1991), who reported that the number of plant breeding graduates declined slightly from 1980 to 1989. Hess (1989) concluded that the decline had bottomed out by 1989, and that some recovery had begun. Our survey results support the conclusion of Hess, showing a 2% increase in the number of plant breeding graduates from 1995 to 2000. Unlike the previous survey results, we found the number of domestic plant breeding graduates increased slightly while the number of international plant breeding graduates decreased.

General Comments from Survey

Telephone conversations with plant breeders revealed problems that were not obvious from the data collected in this survey. First, many plant breeders were operating with fewer resources than in the past. To increase their funding opportunities, breeders were moving to other crop species, or going into more basic areas of research. The reduced state funding and increased pursuit of other funding sources had the effect of reducing the amount of plant breeding work done on each crop. Second, many plant breeders were planning to retire in the next few years and thought that they would not be replaced, or if they were replaced, it would be with someone other than a plant breeder.

Comments received from public plant breeders were generally pessimistic about the future of breeding programs in land-grant universities. Some public plant breeders were well funded, especially those working on self-pollinated crops (such as small grains) or clonally propagated crops (such as sweet-potato, *Ipomoea batatas* var. *batatas*), where there was little private plant breeding effort. However, most plant breeders were not well funded and did not envision a change in the situation in the near future. The shift in government funding away from plant breeding and toward other areas such as molecular genetics has encouraged universities to...
change emphasis in their crop research programs as plant breeders leave or retire. That shift can be seen at universities such as the University of California-Davis, which was formerly a leader in plant breeding but now has a much larger emphasis on molecular genetics.

The number of universities that have maintained a critical mass in plant breeding has been reduced almost to the point of having just one in each region of the USA. Also, some of the university programs, such as the University of Wisconsin-Madison, depend heavily on USDA-ARS scientists who have appointments as departmental faculty, but are supported by the federal government. Therefore, changes in federal support would have a large impact on some departments with graduate training in plant breeding. Universities often require at least six students to be registered for a class before it can be taught. Thus, universities that train fewer than six graduate students per year (all but the top eight universities) may have difficulty in offering their plant breeding courses in the future.

Departments in land-grant universities offer very different emphasis in subdisciplines and commodities, so it would be useful for prospective students to check the specializations of the departments of interest before applying for admission. For example, Washington State University has a large emphasis in the breeding of small grains, the University of Minnesota-St. Paul and Iowa State University are strong in the agronomic crops, and Cornell University, the University of Wisconsin-Madison, and North Carolina State University have programs in both agronomic and horticultural crops.

CONCLUSIONS

As the number of graduate students trained in plant breeding has declined during the last few decades, seed companies have considered hiring students with a B.S. in biological sciences, and then providing training in plant breeding through company programs. This has been acceptable in some cases, particularly floricultural crop breeding, but has not been used at most seed companies. It appears that the downward trend noted in previous surveys has continued to the point where there are only a few universities in each region of the USA that provide training in plant breeding. The USDA advisory committee on agricultural biotechnology has recommended a significant increase in funding for public plant breeders (USDA, 2001). If the USA is interested in continuing the public effort in plant breeding research and graduate student training, sufficient federal and state funding will have to be provided to support at least the current regional centers.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. J.C. Wynne for organizing this survey. This study was funded by the North Carolina Agricultural Research Service.

REFERENCES

The Future of Plant Breeding

Paul Gepts* and Jim Hancock

ABSTRACT

A symposium was hosted 10 to 11 Mar. 2005 at Michigan State University to discuss the future of plant breeding education at public institutions. Plant breeding remains a vibrant, multidisciplinary science characterized by its ability to reinvent itself by absorbing and utilizing novel scientific findings and technical approaches. A contemporary breeding curriculum should include hands-on experience with the inheritance and selection of complex traits in actual plant populations, basic biology of plants (reproductive biology, Mendelian genetics), principles of quantitative genetics and selection theory, principles and practice of plant breeding and related sciences such as genomics, applied statistics, experimental design, and pest sciences. Plant breeding education should also comprise several professional skills, including knowledge of other languages, business management, and intellectual property rights. The private sector should play an increased role in the plant breeding. There is also a need for alternative types of training in plant breeding geared toward working breeders and farmers. Additional support for plant breeding education programs may come from the private sector and the federal government. With regard to specialty crops, increased support for research and education programs may come from a focus on the unique features of these crops. Finally, it is important to cultivate public awareness of the accomplishments of plant breeding.

There has been increasing concern around the world about who will educate plant breeders in the future. To date, much of the M.S. and Ph.D. education in plant breeding has been provided by large public universities in the USA and other developed countries, yet the number of plant breeders at these institutions is steadily declining (Knight, 2003). Support has also declined dramatically for the International Agricultural Research Centers (IARCs) that have traditionally educated students from both developing and developed countries in plant breeding (Khush, 2006).

The loss of plant breeding programs is of great concern to both our domestic plant breeding industry and the international community. The bottom line is that we must sustain a critical mass of applied geneticists and plant breeders at public institutions if we are to maintain our education programs in plant breeding. A symposium, hosted 10 to 11 Mar. 2005 by the Plant Breeding and Genetics Group at Michigan State University, discussed this critical issue, along with the overall education of plant breeders (Hancock, 2006). Several segments of the entire plant breeding community were represented, including private breeding programs, major commodity groups, international training centers, and university breeders.

This paper reviews the major themes discussed during this symposium including (i) defining plant breeding, (ii) describing plant breeding education and employment, (iii) designing a contemporary education in plant breeding, (iv) supporting plant breeding education programs, (v) addressing the critical needs of breeding specialty and subsistence crops, and (vi) promoting awareness of plant breeding.

Define Plant Breeding

Plant breeding is an applied, multidisciplinary science. It is the application of genetic principles and practices associated with the development of cultivars more suited to the needs of humans than the ability to survive in the wild; it uses knowledge from agronomy, botany, genetics, cytogenetics, molecular genetics, physiology, entomology, biochemistry, and statistics (Schlegel, 2003). Of particular importance is the ability to transfer, in addition to major genes, large suites of genes conditioning quantitative traits such as productivity and other traits of interest to humans. The ultimate outcome of plant breeding is mainly improved cultivars. Therefore, plant breeding is primarily an organismal science even though it is eminently suited to translate information at the molecular level (DNA sequences, protein products) into economically important phenotypes. The traditional definition of a plant breeder includes only those scientists who develop new cultivars and improved germplasm; however, many feel this definition should be expanded to include scientists who contribute to crop improvement through breeding research (Ransom et al., 2006).

As a science, plant breeding started soon after the rediscovery of Mendel’s Laws at the beginning of the 20th century. Since then, plant breeding has evolved by absorbing approaches from different areas of science, allowing breeders to increase their efficiency and exploit genetic resources more thoroughly. Over the years, it has put to productive use progress in crop evolution, population and quantitative genetics, statistical genetics and biometry, molecular biology, and genomics. Thus, plant breeding has remained a vibrant science, with continued success in developing and deploying new cultivars on a worldwide basis. On average, around 50% of productivity increases can be attributed to genetic improvement (Fehr, 1984).

Plant Breeding Education and Employment

During the last decade, several surveys have assessed levels of education and employment in plant breeding (Frey, 1996; Traxler et al., 2005; Guner and Wehner, 2003). Important conclusions have come from these...
studies (Baenziger, 2006; Bliss, 2006): (i) A large majority of plant breeders now work in the private (65–75%) vs. public sector (25–35%). [The Frey (1996) survey indicated a general downward trend in public plant breeding employment with a loss in the 1990–1994 period of 2.5 scientist–year (SY) per year, whereas in the private sector there was an increase in the same period of 32 SY per year. Thus, plant breeding employment has increased overall, but has been accompanied by a shift from the public to the private sector.]. (ii) Plant breeders are predominantly occupied with agronomic (75%) vs. horticultural crops (25%). (iii) Plant breeding students are evenly divided between U.S. and foreign citizens. (iv) Developing countries are increasingly important in plant breeding education and research. (v) Private sector breeders are employed to a large extent (80%) in cultivar development, with the remainder distributed about equally between germplasm enhancement and plant breeding research. In the USDA, the proportions are 12, 48, and 40%, respectively, whereas in the state agricultural experiment stations they are 41, 29, and 30%, respectively.

The number of individuals obtaining degrees in plant breeding is relatively stable in recent years. Guner and Wehner (2003) found that 770 plant breeding degrees were awarded in the USA in the period from 1995 to 2000, with no detectable downward trend. This is an annual average of 154 plant breeding graduates, distributed about equally between domestic and international students. Bliss (2006) estimated that the USA requires 45 to 110 graduates each year, assuming turnover rates of 2 to 5% in the current pool of 2200 plant breeders. Hence, the job market for plant breeders remains quite strong with salaries comparing favorably with other occupations in the agricultural sciences. No full scale surveys of plant breeding exist for the rest of the world, although the demand for new breeders appears to be low in developing countries (Guimarães and Kueneman, 2006).

While the number of plant breeding graduates is currently adequate to meet U.S. and perhaps world needs, the university education programs that will produce the breeders of the future are themselves at risk. The number of plant breeders employed by agricultural experiment stations dropped by 21% just from 1994 to 2001, leaving only 420 plant breeders associated with universities (Traxler et al., 2005). Guner and Wehner (2003) and Ransom et al. (2006) posit that there are now only a handful of institutions in the USA with the critical mass of faculty necessary to offer a complete plant breeding curriculum. The loss of even a few of these programs will mean insufficient breeders for future generations. Thus, we are in a paradox of a strong demand for plant breeding graduates to fill well-paying, scientific jobs, yet public funding and education in this area are declining and will ultimately result in shortages of plant breeders.

Several negative factors weaken the strength of plant breeding programs at public institutions (Baenziger, 2006; Guner and Wehner, 2003). As plant breeders retire, they are replaced by scientists involved in more basic genetic studies. This shift is fueled by the perception that private sector breeding efforts are adequate to meet cultivar needs. Cuts in University resources have also led to reduced support of field programs. This has pushed the current crop of public plant breeders to shift their activities toward fundamental/basic studies that can be supported by federal grants and the private sector. These negative factors have reduced the number of plant breeding faculty available in the future to teach the range of courses needed to support a plant breeding curriculum.

**Other Trends Affecting Plant Breeding**

Plant breeding is affected by increasing globalization in two major ways: (i) an increasing commercialization of agricultural products with its additional export opportunities, but also (ii) increased competition from other regions with lower production costs (Morris et al., 2006). Furthermore, new trade rules and the award (or elimination) of subsidies can rapidly alter the competitive climate among countries. Such changes can dramatically alter the value of individual commodities at the international level, increasing or decreasing the importance of plant breeding programs over short time periods (Morris et al., 2006).

The concept of ownership of biodiversity has changed radically during the last 25 yr. Whereas until 1980, genetic diversity was considered the common heritage of humankind, it is increasingly being subjected to laws and treaties that allow institutions and companies employing plant breeders to claim ownership on cultivars (Gepts, 2004). The full effect of this change on plant breeding innovation, germplasm accessibility, and exchange remains to be determined, but it is limiting the exchange of plant materials among breeding programs because of the increase in ownership claims such as patents, plant variety protection certificates, and material transfer agreements.

Overall, public research and development (R&D) investment has decreased but private R&D investment has increased (Morris et al., 2006), which mirrors the employment trends mentioned earlier. Public research programs have also qualitatively changed their focus, as funding has shifted from applied field experiments to genomics and molecular biology. A direct offshoot of these more basic approaches, genetic engineering is presented erroneously as a speedier, more precise plant breeding alternative, whereas genetic engineering is simply another way of generating genetic diversity at the onset of a plant breeding cycle (Gepts, 2002). In short, improved cultivars are still generated through conventional approaches. That is why most agricultural biotechnology companies have purchased breeding companies.

Pursuing genetic engineering as an alternative rather than as a complement to plant breeding has large consequences on choices about R&D investment, education, and employment. This is increasingly apparent in many developing countries, where plant breeders are replaced by biotechnologists (Morris et al., 2006; Guimarães and Kueneman, 2006). These allocation changes will not enhance agricultural production unless a critical mass of
plant breeders remains to translate the technologies into new cultivars.

**Designing a Contemporary Education in Plant Breeding**

The key to plant breeding education is the exposure to inheritance and selection of complex traits in actual plant populations (Ransom et al., 2006). To fully gauge the challenge of plant breeding, students should witness first hand the effect of selection on the inheritance of quantitative traits as well as the influence of the environment.

The essential categories of scientific knowledge needed by all plant breeders include (i) principles and practice of plant breeding; (ii) Mendelian/transmission genetics; (iii) applied statistics and experimental design; (iv) quantitative genetics and selection theory; (v) production principles and practices for agronomic, horticultural, and tree crops; (vi) pest sciences (plant pathology, entomology, and weed science); (vii) applied genomics, including marker-assisted selection; and (viii) plant reproductive biology (Bliss, 2006; Ransom et al., 2006). Additional desirable areas of lesser importance are (i) biotechnology (tissue culture, genetic transformation), (ii) evolutionary and population genetics, (iii) crop physiology and plant biochemistry, and (iv) cytogenetics. In designing a new curriculum, efforts should be made to integrate newer areas into existing courses, rather than just increasing the number of courses. Plant breeding education should also comprise several professional skills. These include (i) knowledge of other languages besides English, (ii) business management: human resources, priority setting, organizational, and budget, (iii) intellectual property rights, and (iv) leadership and teamwork.

Many university-educated plant breeders will be employed in private industry, requiring greater involvement of private sector plant breeders in public breeding educational programs. New models of collaboration are needed between universities and industry to ensure that new Ph.D. students comfortably fit into the private sector and company breeders can reenter the university as professors (Baenziger, 2006). The private sector can play an increased role in the plant breeding curriculum through (i) participation in thesis and curriculum committees, (ii) guest lectures, seminars, Q&A sessions, (iii) internships with private company breeders, (iv) endowment of professorships or breeding programs, (v) provision of scholarships, and (vi) provision of grants for research, study, or conferences. Novel collaborations could also be forged between universities and plant breeding research institutes, such as the centers of the Consultative Group on International Agricultural Research (CGIAR). The universities could provide basic education in plant breeding, whereas research institutes would provide practical and field breeding experience.

Industry demands individuals with degrees at all levels, B.S., M.S., and Ph.D. Individuals with these different degrees are involved in different types of activity, with lower degrees more involved in program support and higher degrees in technical support and especially cultivar development. Several other types of education or training in plant breeding are needed: (i) M.S. or M.B.A. degree geared to working breeders; (ii) updates for plant breeders on new technologies such as applied genomics and marker-assisted selection, as well as intellectual property rights and biosafety; (iii) nondegree course work training for agronomic professionals in topics such as breeding practices, basic genetics, and selection theory; and (iv) participatory plant breeding training for farmers (parabreeders) in developing countries with an important subsistence sector.

**Supporting Plant Breeding Education Programs**

To maintain a critical mass of graduate education programs, new funding sources are needed to support the existing programs (Ransom et al., 2006; Terpstra et al., 2006). The private sector would greatly benefit from more generously supporting the public education programs on which they depend for future plant breeders (Bliss, 2006). One approach is to establish a national fund for graduate fellowships and undergraduate internships in plant breeding that is supported by contributions from many different corporations. Individual companies could also increase the amount of grants they award for cooperative research that directly benefits them. For example, a public breeder is given a grant to evaluate native germplasm that will be used later by an industry breeder for cultivar development, or to identify novel quantitative trait loci for use by the company in marker-assisted breeding.

The federal government also needs to increase support for plant breeding education (Ransom et al., 2006; Terpstra et al., 2006). Federal funds should be used to directly support the elite education programs, stimulate educational linkages between universities, and encourage public–private sector collaborations. The only national program that we are aware of which can be accessed to support this type of education is the most recent National Needs Fellowship Program of the USDA. Many students interested in plant breeding are supported on federal competitive grants, but these funds are generally awarded for very basic projects that are far downstream from actual cultivar improvement. In reality, the students supported by these programs generally get little practical plant breeding experience, as they are committed to finishing projects at the bench.

**Addressing the Critical Needs of Breeding of Specialty Crops**

Specialty and subsistence crops are characterized by a limited acreage and/or low gross income (Weebadde and Mensah, 2006). However, they fulfill niche markets important for local economies or address specific human or societal needs. Support for subsistence crops presents a particularly thorny problem, as little profit is gained from their development. Such crops are often most important in countries with limited resources. Because of the low gross income, they are generally neglected by the private sector as lacking sufficient potential for eco-
momic return. As a consequence, the responsibility for research on these crops often falls to public universities and CGIAR centers whose funding base continues to erode. For example, the number of public sector breeders working on fruit and vegetable crops in the USA has declined by 43% over the last decade (Traxler et al., 2005).

To achieve a critical mass of specialty and subsistence crop breeders, there is an obvious need to develop institutional alliances, networks, or partnerships, in the same or most likely different institutions. Each member will have their specialty within the network, including plant breeding. Specialty and subsistence crop breeders often work on more than one crop. Part of the networking may involve nontraditional alliances such as seed saver organizations or small companies. Networks may also involve local farmers or parabreeders who participate in germplasm selection with research centers or public universities (similar to participatory plant breeding). Separate local seed production microenterprises could be set up as well.

New funding opportunities must be developed to promote collaborations among different states and countries so that specialty and subsistence crop breeders work together instead of competing for funds. Federal and international grants could be directed to the special features of specialty crops rather than just the crop itself, such as close taxonomic relationships with major crops (e.g., synten or small genome size), interesting features such as a role as a medicinal plant, a functional food, a novelty food (e.g., tropical fruit), and as alternative use (e.g., biofuel, alternative source of an industrial product). More patenting and licensing of specialty crops should also be undertaken to stimulate entrepreneurship in the private sector and provide means to collect royalties from licensing agreements. Plant breeding education could benefit significantly by improving the funding base for specialty crop breeding, as the breeders of these crops will most likely be at public institutions.

Promoting Awareness of Plant Breeding

Perhaps most critical to the future of plant breeding is documentation and publication of its key role in the tremendous increases in productivity from the 20th century onward, as documented in the introduction. Urban populations in particular need to be convinced that plant breeding plays a key role in societal well-being, through food security and high quality vegetables and fruits. Raising the awareness of plant breeding by students in urban schools at various levels, from high school to bachelors level students, is a priority. Projects can be developed with federal funding to illustrate the concepts of inheritance and selection not only of qualitative but also quantitative traits using computer models and short-cycle plants (e.g., Wisconsin Fast Plants, www.fastplants.org, verified 14 Mar. 2006). Plant breeders should speak with a unified voice to capture the attention of the policymakers and the imagination of the public.

Given the expanded role of the private sector in plant breeding, industry organizations should be enlisted to promote plant breeding and especially, the education of plant breeders. Among these organizations are the American Seed Trade Organization (ASTA, www.amseed.com, verified 14 Mar. 2006), the National Council of Commercial Plant Breeders (NCCPB, www.nccpb.org, verified 14 Mar. 2006), and many commodity organizations.

Plant breeders and their students should also more fully interact with grassroots movements involved in genetic conservation such as Slow Food (www.slowfood.com, verified 14 Mar. 2006), the Seed Savers Exchange (www.seedsavers.org/Home.asp, verified 14 Mar. 2006), Native Seeds Search (www.nativeseeds.org, verified 14 Mar. 2006), Southern Exposure Seed Exchange (www.southernexposure.com, verified 14 Mar. 2006), the Appalachian Heirloom Seed Conservancy (Richmond, KY), the Northern Plains Sustainable Agriculture Society (www.npsas.org, verified 14 Mar. 2006), and more local organizations.

Another source of support to tap is alumni, who can speak firsthand of the accomplishments of plant breeding and can provide advice on how to improve plant breeding education. Plant breeders should work with campus development officers to maintain contact with alumni. In addition, fundraising campaigns should be established to appeal to constituencies or stakeholders to establish endowments that will support education and field-based research.

A national plant breeding coordinating committee needs to be assembled that will establish a roadmap for the future of plant breeding. Ann Marie Thro, CSREES National Program Leader, has begun organizing such a group (www.csrees.usda.gov/nea/plants/in_focus/pbgg_if_multistate.html, verified 27 Mar. 2006). This committee should be composed of plant breeders from both the private and public sector, along with representatives of the major commodity groups and funding agencies. We hope that this group will find the outlined symposium proceedings useful in charting the future of plant breeding. A holistic strategy needs to be developed to strengthen our plant breeding capacity and it is critical that this committee finds a way to communicate the importance of plant breeding to the lay public.

In conclusion, we favor a multipronged approach to reinvigorate the plant breeding education and research enterprise. After a first century of outstanding successes, plant breeding remains a vital science which constantly renews itself to absorb new technical and scientific advances. Therefore, it remains an essential part of crop improvement, especially in light of increased demands for food, feed, and fiber resulting from an ever-increasing world population in this second century of plant breeding.

REFERENCES

Fehr, W.R. (ed.). 1984. Genetic contributions to yield gains of five major crop plants. CSSA Spec. Publ. 7. ASA and CSSA. Madison, WI.
Frey, K. 1996. National plant breeding study: I. Human and financial resources devoted to plant breeding research and development in