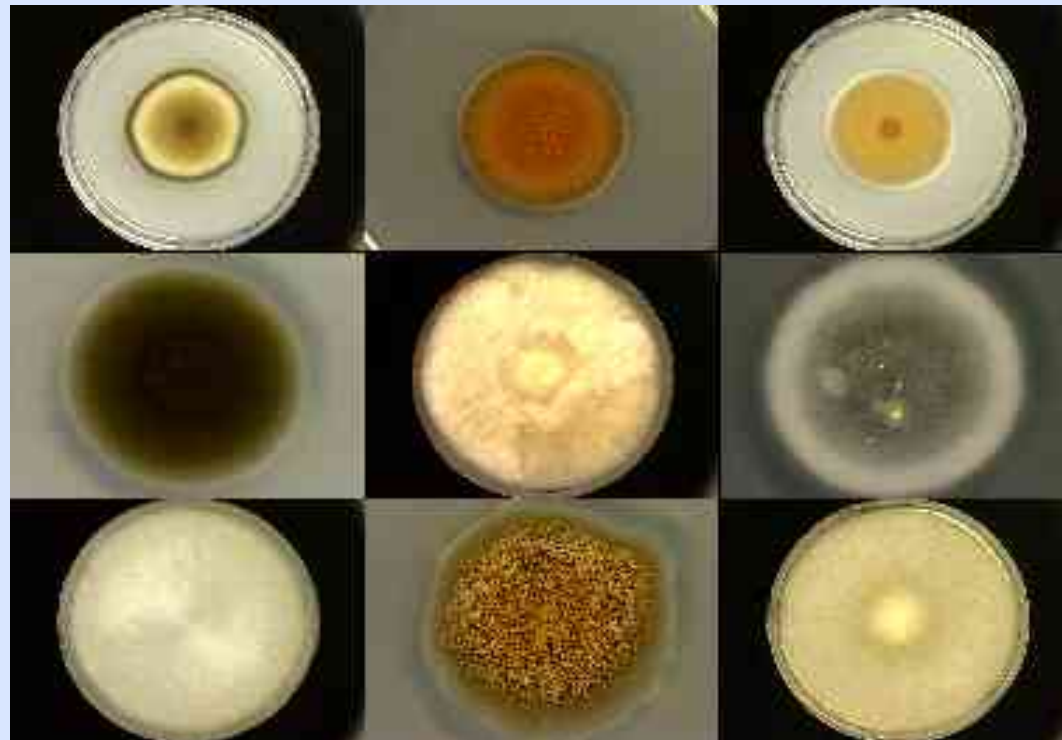


# Michael Stelzer

## Eine Einführung in die Hierarchische Clusteranalyse (HCA)



# Inhalt

---

**1. Problemstellung**

**2. Agglomerationsmethoden  
und Algorithmen**

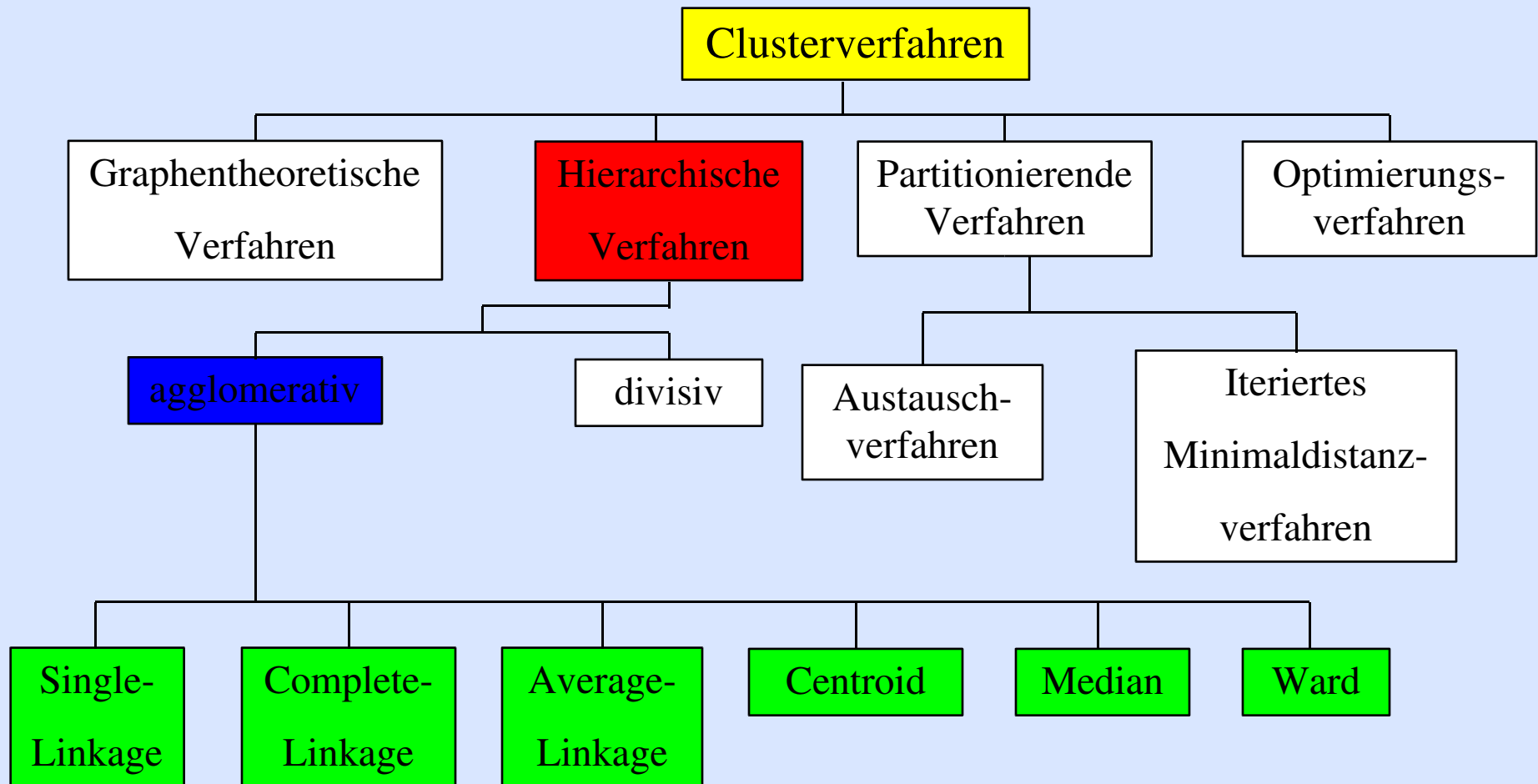
**3. Anwendungsbeispiele**

**4. Zusammenfassung**

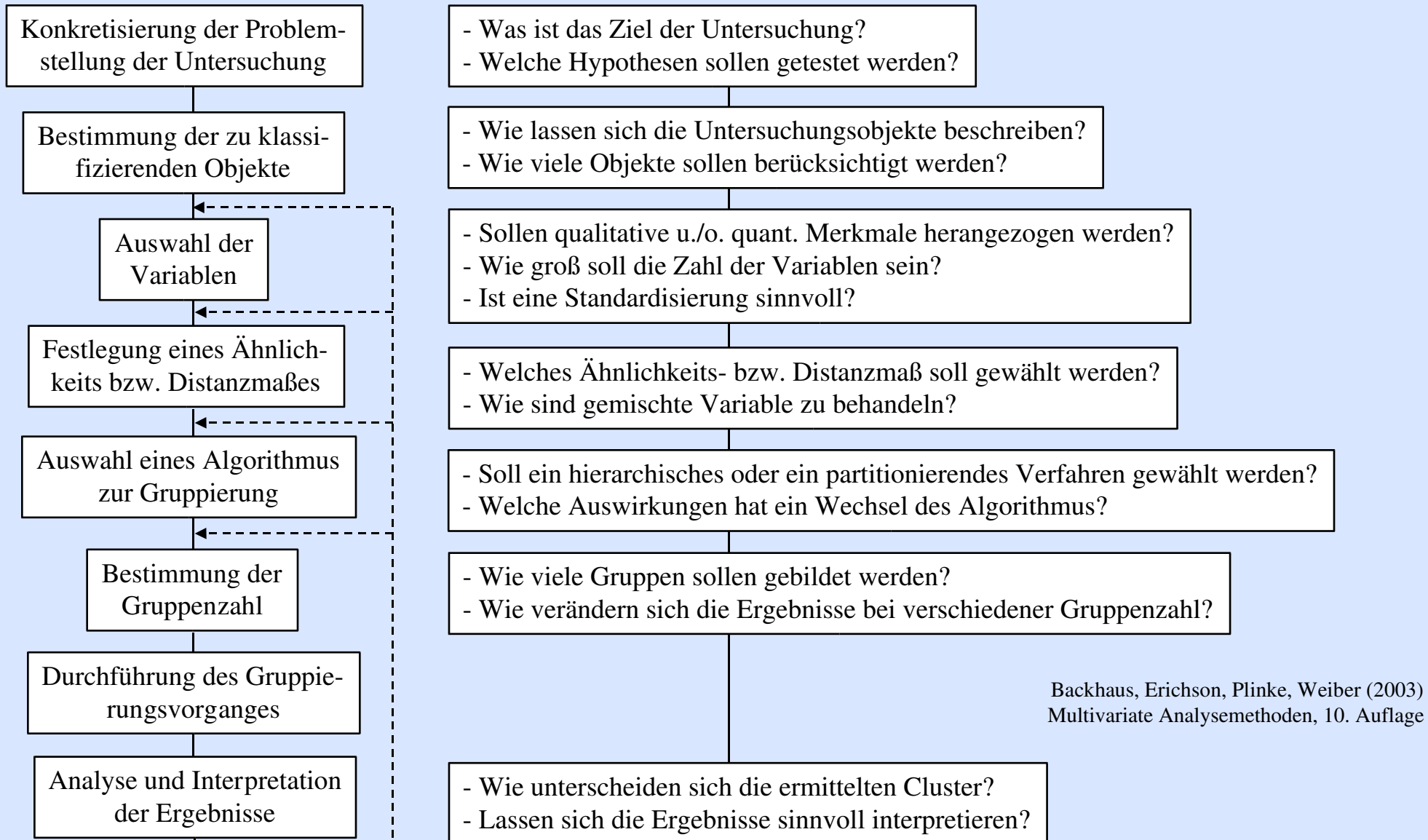
# Problemstellung

- **Cluster:** dt.: Traube, Haufen
- **heuristisches Verfahren:** systematische Klassifizierung von Beobachtungen (Personen, Autos, CD's...)  
heterogene Gesamtheit  $\rightarrow$  homogene Teilmenge von Objekten
- **Anwendung:** Sozial- + Wirtschaftswissenschaft, Marktforschung, Biologie
- **Rohdatenmatrix**  $\rightarrow$  Distanz-/Ähnlichkeitsmatrix
- **Ähnlichkeit/Unähnlichkeit** durch Merkmale wie z.B. Alter, Haarfarbe definiert
  - Unähnlichkeit = **Distanz** (distance): absoluter Abstand
  - Ähnlichkeit = **Proximität** (similarity): primärer Ähnlichkeitsaspekt
- **Dendrogramm:** binär verzweigter Verwandtschaftsstammbaum  $\rightarrow$  Hierarchie
- **Clusterzahl (CZ):** Handhabbarkeit (geringe CZ)    Homogenitätsanforderung (große CZ)
- Unterscheidung nach **Zahl** verwendeter **Variablen:**
  - monothetische Verfahren: nur 1 Variable herangezogen
  - polythetische Verfahren: mehrere Variablen eingesetzt

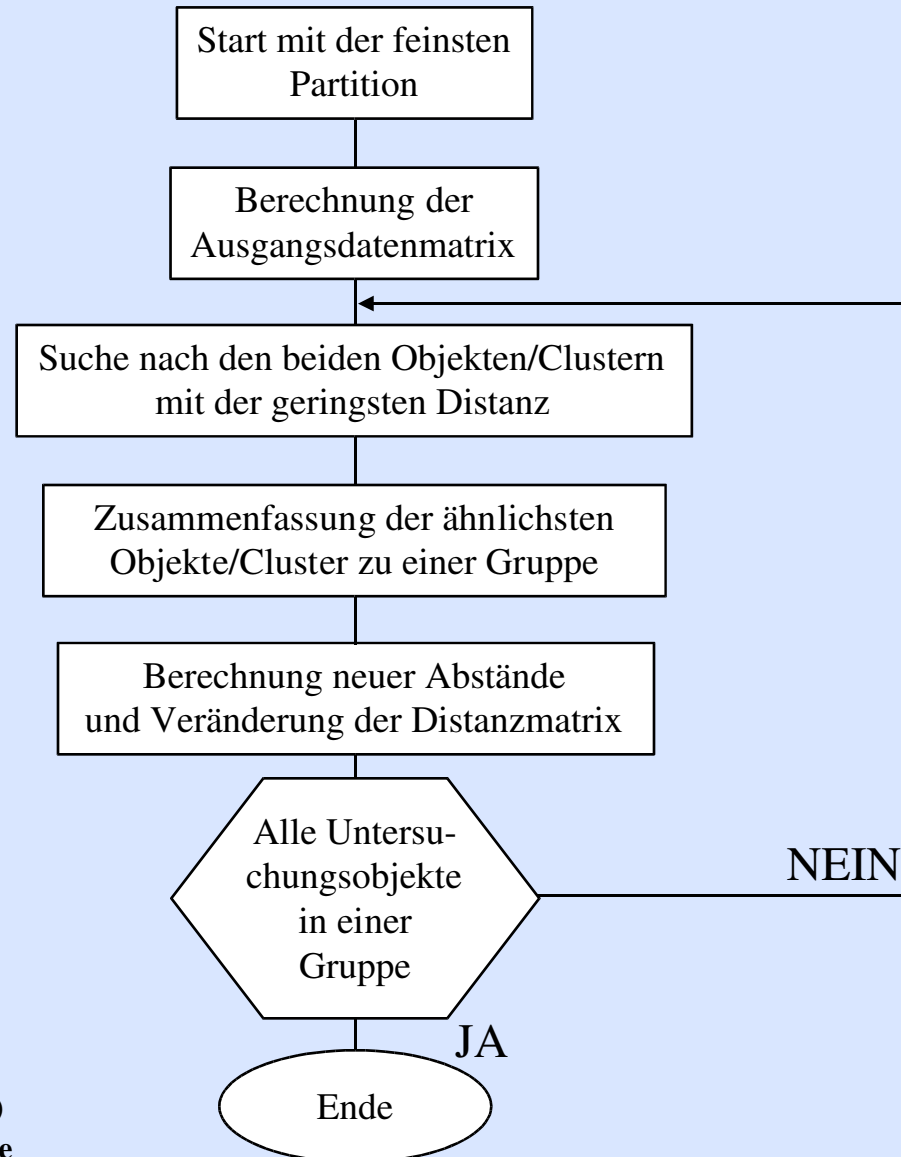
# Überblick über ausgewählte Cluster-Algorithmen



# Ablaufschritte und Entscheidungsprobleme der Clusteranalyse



# Ablaufschritte der agglomerativen hierarchischen Clusterverfahren



# Charakterisierung agglomerativer Clusterverfahren (WinSTAT®)

## - Einfache Verbindung (Single/Simple Linkage):

Distanz zweier Gruppen = Distanz beider am dichtesten benachbarten Objekte  
→ kontrahierend: viele kleine, wenige große Gruppen + Kettenbildung + Ausreißer

## - Komplette Verbindung (Complete Linkage):

Distanz zweier Gruppen = Distanz beider am weitesten entfernten Objekte  
→ dilatierend: kleine, gleich große kompakte Cluster + anfällig für Ausreißer

## - Durchschnittsverbindung (Average Linkage):

Distanz zweier Gruppen = mittlere Distanz aller mögl. Verbindungen der Objekte  
→ konservativ + Tendenz zur Verbindung Cluster kleiner Varianz → gleiche Var.

## - Zentroidverbindung (Centroid Linkage):

Distanz zweier Gruppen = Distanz der Gruppenschwerpunkte  
→ konservativ + robust gegenüber Ausreißern

## - Ward-Methode (Incremental Sums of Squares) (1963):

Distanz zweier Gruppen ~ Änderung der  $\sum$  der Distanzen innerhalb der Gruppen, die sich bei Zusammenschluss ergeben würde  
→ konservativ + homogene, gleich große Cluster + anfällig für Ausreißer

# Charakterisierung agglomerativer Clusterverfahren (XLstat<sup>©</sup>)

## - **Unweighted Pair Group Method with Arithmetic Mean:**

Unähnlichkeit zwischen zwei Gruppen ist Mittel der Unähnlichkeiten zwischen den Objekten beider Gruppen → guter Kompromiss zwischen Simple u. Complete Linkage + faire Darstellung der Datenraummöglichkeiten

## - **Weighted Pair Group Method with Arithmetic Mean:**

mittlere Unähnlichkeit zwischen Objekten zweier Gruppen ist  $\Sigma$ gewichteter Unähnlichkeiten → beiden Gruppen gleiche Gewichtung zugewiesen + ehrliche Repräsentation des Datensatzes

## - **Flexible Linkage:**

$\beta$ -Parameter  $[-1,+1]$ ;  $\beta = 0$  → WPGMA;  $\beta$  nahe 1 → kettenartige Cluster;  
 $\beta < 0$  → Ausdehnung des Datensatzes

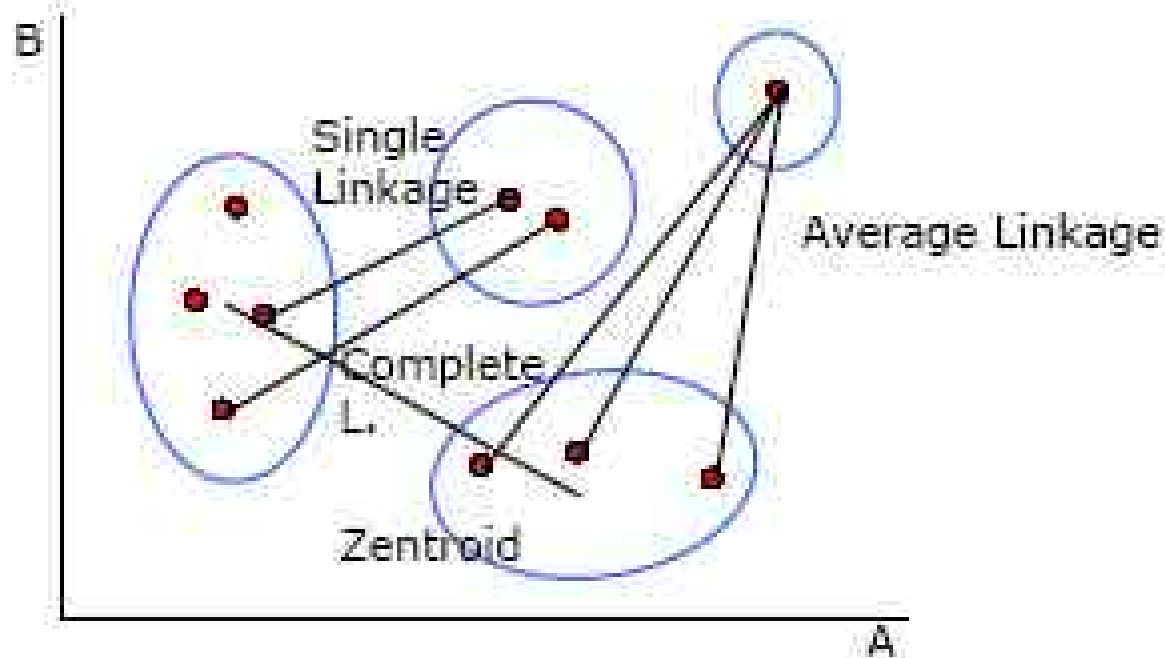
## - **Strong Linkage:**

verwendet mittlere Distanzen innerhalb jeder Gruppe + zwischen allen Gruppen  
→ sehr kompakte Cluster

# Vergleich agglomerativer Clusterverfahren

## Beispiel

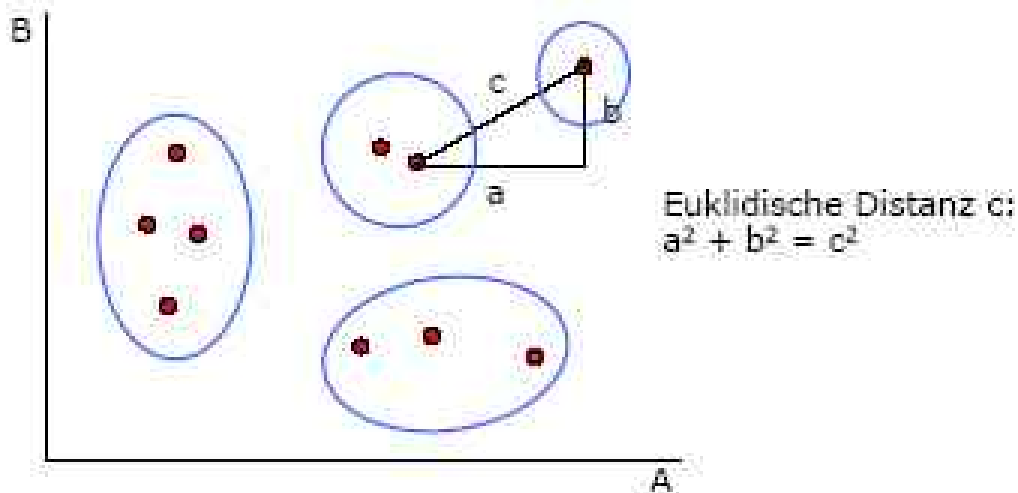
- Zehn Fälle, zwei stetige Merkmale A/B



# Die Euklidische Distanz

## Beispiel

- o Zehn Fälle, zwei stetige Merkmale A/B



### Quadrierte ED:

große Differenzwerte stärker berücksichtigt als kleine Werte → Gewichtung

$$P \text{ Merkmale: } d_{ii'} = \left[ \sum_{i=1}^p (X_{ij} - X_{i'j})^2 \right]^{1/2}$$

Beispiel für 2 Fälle A und B, 7 Merkmale:

A: 5 7 8 1 3 2 5

B: 9 5 8 2 7 8 2

$$d_{ab} = \sqrt{(5-9)^2 + (7-5)^2 + \dots + (5-2)^2} = 9,055$$

# Distanzberechnung bei ausgewählten agglomerativen Verfahren

| Verfahren                     | Konstante                |                          |                                  |      | Distanzberechnung (D(R;P + Q)) nach Gleichung X):  |
|-------------------------------|--------------------------|--------------------------|----------------------------------|------|--|
|                               | A                        | B                        | E                                | G    |  |
| Single Linkage                | 0,5                      | 0,5                      | 0                                | -0,5 | $0,5 \cdot \{D(R,P) + D(R,Q) -  D(R,P) - D(R,Q) \}$  |
| Complete Linkage              | 0,5                      | 0,5                      | 0                                | 0,5  | $0,5 \cdot \{D(R,P) + D(R,Q) +  D(R,P) - D(R,Q) \}$  |
| Average Linkage (ungewichtet) | 0,5                      | 0,5                      | 0                                | 0    | $0,5 \cdot \{D(R,P) + D(R,Q)\}$  |
| Average Linkage (gewichtet)   | $\frac{NP}{NP+NQ}$       | $\frac{NQ}{NP+NQ}$       | 0                                | 0    | $\frac{1}{NP+NQ} \{NP \cdot D(R,P) + NQ \cdot D(R,Q)\}$  |
| Centroid                      | $\frac{NP}{NP+NQ}$       | $\frac{NQ}{NP+NQ}$       | $-\frac{NP \cdot NQ}{(NP+NQ)^2}$ | 0    | $\frac{1}{NP+NQ} \{NP \cdot D(R,P) + NQ \cdot D(R,Q)\} - \frac{NP \cdot NQ}{(NP+NQ)^2} \cdot D(P,Q)$           |
| Median                        | 0,5                      | 0,5                      | -0,25                            | 0    | $0,5 (D(R,P) + D(R,Q)) - 0,25 \cdot D(P,Q)$  |
| Ward                          | $\frac{NR+NP}{NR+NP+NQ}$ | $\frac{NR+NQ}{NR+NP+NQ}$ | $-\frac{NR}{NR+NP+NQ}$           | 0    | $\frac{NR+NP}{NR+NP+NQ} \cdot D(R,P) + \frac{NR+NQ}{NR+NP+NQ} \cdot D(R,Q) - \frac{NR}{NR+NP+NQ} \cdot D(P,Q)$ |

Gleichung X:  $D(R,P + Q) = A \cdot D(R,P) + B \cdot D(R,Q) + E \cdot D(P,Q) + G \cdot |D(R,P) - D(R,Q)|$

NR: Zahl der Objekte in Gruppe R

NP: Zahl der Objekte in Gruppe P

NQ: Zahl der Objekte in Gruppe Q

D(R,P): Distanz zw. Gruppen R + P

D(R,Q): Distanz zw. Gruppen R + Q

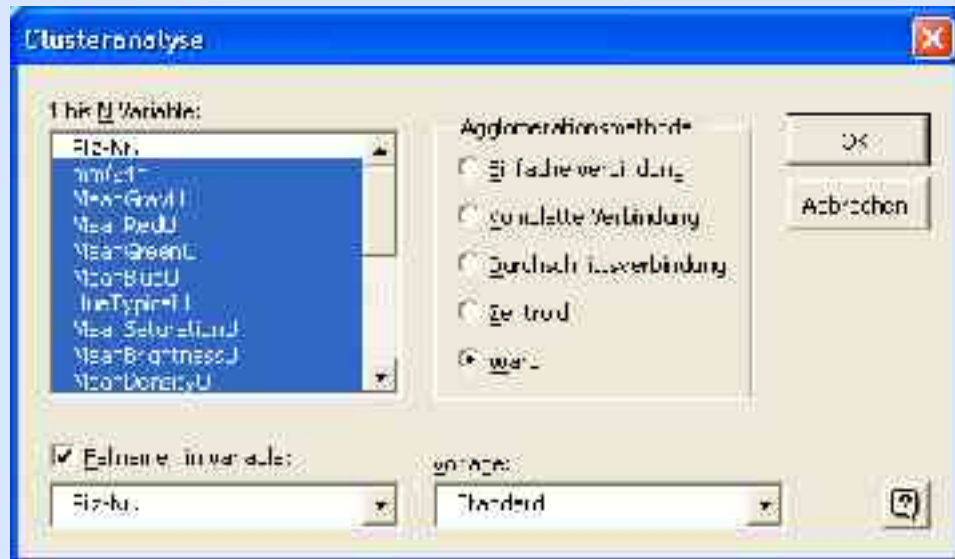
D(P,Q): Distanz zw. Gruppen P + Q

# Programmpakete

---

- **XLstat**<sup>®</sup>: Vers. 5.1, Addinsoft, Dr. Thierry Fahmy, Paris, Frankreich (Windows)
- **WinSTAT**<sup>®</sup>: Vers. 1999.3, R. Fitch Software, Staufen, Deutschland (Windows)
- **Pirouette**<sup>®</sup>: Vers. 3.11, Infometrix Software Inc., Dr. E. Riverside, U.S.A. (Windows)
- **SPSS**<sup>®</sup>: Vers. 13, SPSS Inc., U.S.A. (Windows + Linux?)
- **R**: Vers. 2.0.1 (15.11.2004) The R Foundation for Statistical Computing, R. Gentleman + R. Ihaka *R&R* of the Statistics Department Univ. of Auckland (Windows + Linux)
- **GelCompar**<sup>®</sup> **II**: Applied-Maths B/BA, Belgien (Windows)
- **Unistat**: Vers. 5, Additive GmbH, Deutschland (Windows)
- ...
- ...

# Vorgehensweise in WinSTAT® (I)



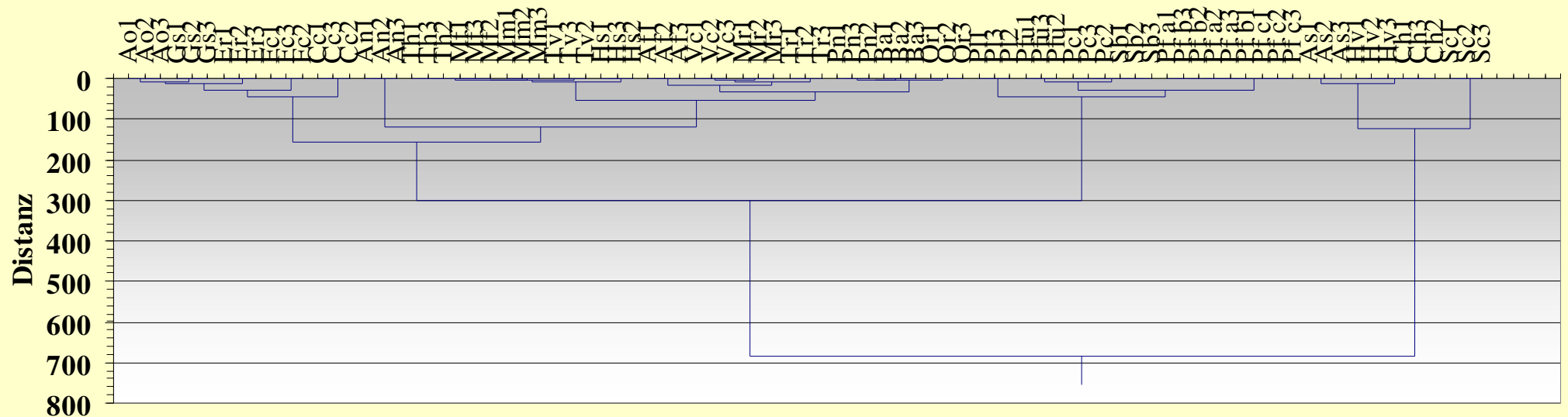
Einstellungen Dialog *Clusteranalyse*

(1)

Einstellungen Dialog *Clustertrennung*

(3)

(2) Ergebnisse Blatt *Cluster*



# Vorgehensweise in WinSTAT® (II)

## Clusteranalyse

### Meßvariable:

mm/24h

MeanGrayU

MeanRedU

MeanGreenU

MeanBlueU

HueTypicalU

MeanSaturationU

MeanBrightnessU

MeanDensityU

MeanGrayR

MeanRedR

MeanGreenR

MeanBlueR

HueTypicalR

MeanSaturationR

GrayVariationR

MeanBrightnessR

BrightVariationR

MeanDensityR

Agglomerations-  
methode: Ward

verbinde

mit

Schritt

Cluster 1

Größe 1

Cluster 2

Größe 2

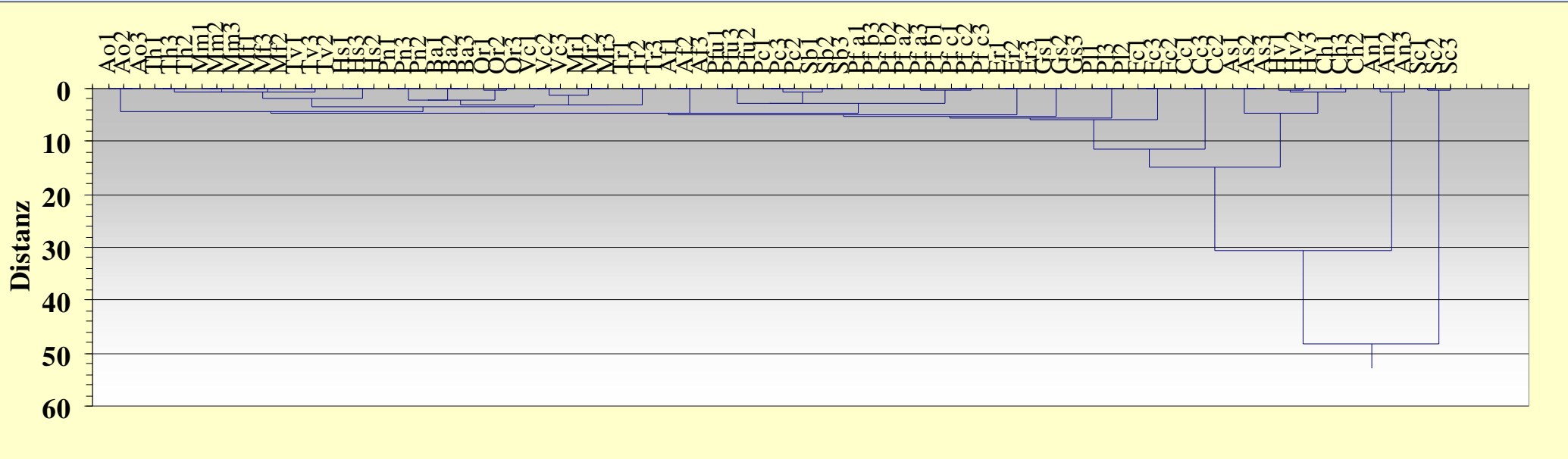
Distanz

|    |       |   |       |   |      |
|----|-------|---|-------|---|------|
| 1  | Gs2   | 1 | Gs3   | 1 | 0    |
| 2  | Cc1   | 1 | Cc3   | 1 | 0    |
| 3  | Vc1   | 1 | Vc2   | 1 | 0    |
| 4  | Tr1   | 1 | Tr2   | 1 | 0,01 |
| 5  | Mr2   | 1 | Mr3   | 1 | 0,01 |
| 6  | Mm1   | 1 | Mm2   | 1 | 0,01 |
| 7  | Tv1   | 1 | Tv3   | 1 | 0,01 |
| 8  | Af1   | 1 | Af2   | 1 | 0,01 |
| 9  | Pl1   | 1 | Pl3   | 1 | 0,01 |
| 10 | Ao2   | 1 | Ao3   | 1 | 0,01 |
| 11 | Sc1   | 1 | Sc2   | 1 | 0,01 |
| 12 | Pc1   | 1 | Pc3   | 1 | 0,01 |
| 13 | Mm1   | 2 | Mm3   | 1 | 0,01 |
| 14 | Mf1   | 1 | Mf3   | 1 | 0,01 |
| 15 | Th1   | 1 | Th3   | 1 | 0,02 |
| 16 | Pf a1 | 1 | Pf b3 | 1 | 0,02 |
| 17 | Mf1   | 2 | Mf2   | 1 | 0,02 |
| 18 | Ba2   | 1 | Ba3   | 1 | 0,02 |
| 19 | Er1   | 1 | Er2   | 1 | 0,02 |
| 20 | Pf c1 | 1 | Pf c2 | 1 | 0,02 |
| 21 | Ao1   | 1 | Ao2   | 2 | 0,02 |
| 22 | Fc1   | 1 | Fc3   | 1 | 0,02 |
| 23 | As2   | 1 | As3   | 1 | 0,02 |

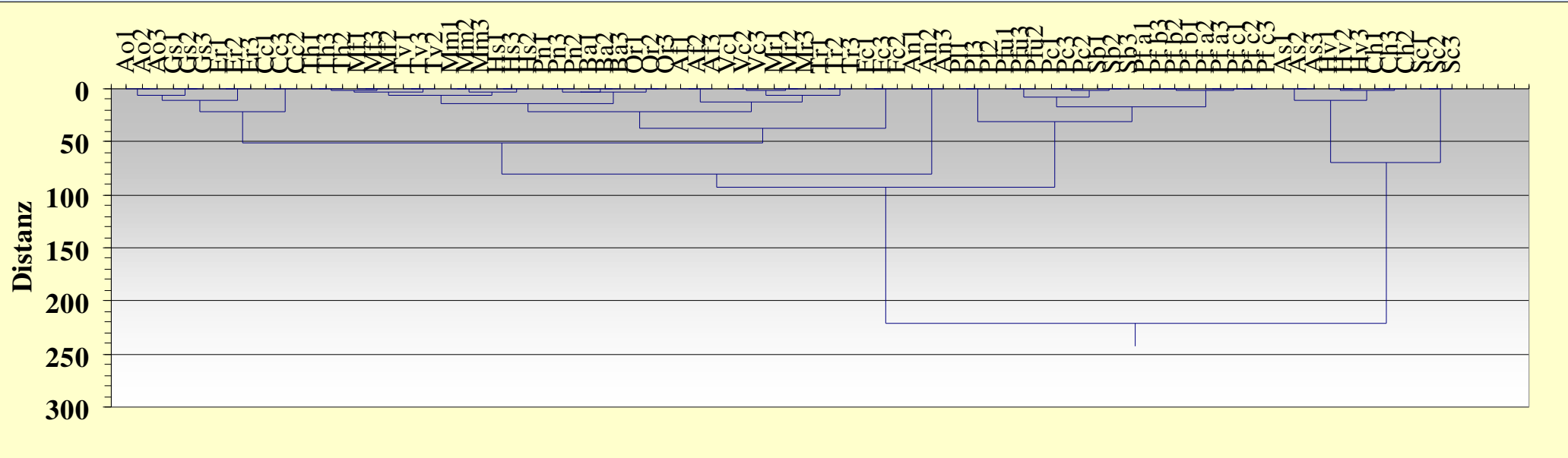
# Vorgehensweise in WinSTAT® (III)

| Schritt | verbinde  |         | mit       |         | Distanz | Schritt | verbinde  |         | mit       |         | Distanz |
|---------|-----------|---------|-----------|---------|---------|---------|-----------|---------|-----------|---------|---------|
|         | Cluster 1 | Größe 1 | Cluster 2 | Größe 2 |         |         | Cluster 1 | Größe 1 | Cluster 2 | Größe 2 |         |
| 27      | Ch1       | 1       | Ch3       | 1       | 0,02    | 57      | Pf a1     | 3       | Pf a2     | 2       | 0,18    |
| 28      | Ba1       | 1       | Ba2       | 2       | 0,02    | 58      | Pf a1     | 5       | Pf b1     | 1       | 0,31    |
| 29      | Gs1       | 1       | Gs2       | 2       | 0,03    | 59      | An1       | 2       | An3       | 1       | 0,39    |
| 30      | An1       | 1       | An2       | 1       | 0,03    | 60      | Pf a1     | 6       | Pf c1     | 3       | 0,82    |
| 31      | Hv2       | 1       | Hv3       | 1       | 0,03    | 61      | Th1       | 3       | Mf1       | 3       | 1,12    |
| 32      | Th1       | 2       | Th2       | 1       | 0,03    | 62      | Hv1       | 3       | Ch1       | 3       | 1,6     |
| 33      | Tv1       | 2       | Tv2       | 1       | 0,03    | 63      | Pc1       | 3       | Sb1       | 3       | 1,68    |
| 34      | Hs1       | 1       | Hs3       | 1       | 0,03    | 64      | Vc1       | 3       | Mr1       | 3       | 2,21    |
| 35      | Sb2       | 1       | Sb3       | 1       | 0,04    | 65      | Th1       | 6       | Mm1       | 3       | 2,71    |
| 36      | Cc1       | 2       | Cc2       | 1       | 0,04    | 66      | Pn1       | 3       | Ba1       | 3       | 3,53    |
| 37      | Sb1       | 1       | Sb2       | 2       | 0,04    | 67      | Th1       | 9       | Tv1       | 3       | 4,15    |
| 38      | Af1       | 2       | Af3       | 1       | 0,04    | 68      | Pn1       | 6       | Or1       | 3       | 4,58    |
| 39      | Tr1       | 2       | Tr3       | 1       | 0,04    | 69      | Vc1       | 6       | Tr1       | 3       | 7,3     |
| 40      | Or2       | 1       | Or3       | 1       | 0,05    | 70      | Th1       | 12      | Hs1       | 3       | 7,32    |
| 41      | Fc1       | 2       | Fc2       | 1       | 0,05    | 71      | Ao1       | 3       | Gs1       | 3       | 8,17    |
| 42      | Pfu1      | 1       | Pfu3      | 1       | 0,05    | 72      | Pfu1      | 3       | Pc1       | 6       | 9,67    |
| 43      | Pn1       | 1       | Pn3       | 1       | 0,05    | 73      | Ao1       | 6       | Er1       | 3       | 13,26   |
| 44      | Pf a1     | 2       | Pf b2     | 1       | 0,05    | 74      | As1       | 3       | Hv1       | 6       | 13,79   |
| 45      | As1       | 1       | As2       | 2       | 0,06    | 75      | Af1       | 3       | Vc1       | 9       | 15,63   |
| 46      | Pf a2     | 1       | Pf a3     | 1       | 0,07    | 76      | Ao1       | 9       | Fc1       | 3       | 26,9    |
| 47      | Pn1       | 2       | Pn2       | 1       | 0,07    | 77      | Pfu1      | 9       | Pf a1     | 9       | 28,55   |
| 48      | Hs1       | 2       | Hs2       | 1       | 0,07    | 78      | Af1       | 12      | Pn1       | 9       | 33,57   |
| 49      | Pc1       | 2       | Pc2       | 1       | 0,08    | 79      | Ao1       | 12      | Cc1       | 3       | 44,61   |
| 50      | Vc1       | 2       | Vc3       | 1       | 0,08    | 80      | Pl1       | 3       | Pfu1      | 18      | 47,25   |
| 51      | Er1       | 2       | Er3       | 1       | 0,09    | 81      | Th1       | 15      | Af1       | 21      | 54,08   |
| 52      | Pfu1      | 2       | Pfu2      | 1       | 0,1     | 82      | An1       | 3       | Th1       | 36      | 118,11  |
| 53      | Ch1       | 2       | Ch2       | 1       | 0,12    | 83      | As1       | 9       | Sc1       | 3       | 124,84  |

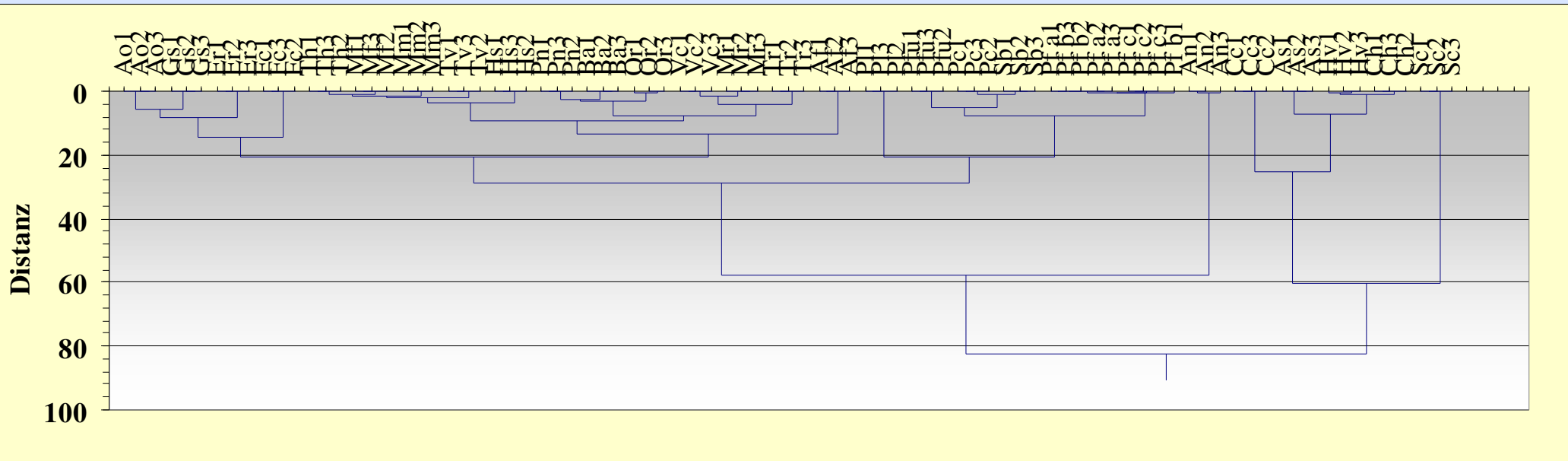
# Einfache Verbindung (Single/Simple Linkage)



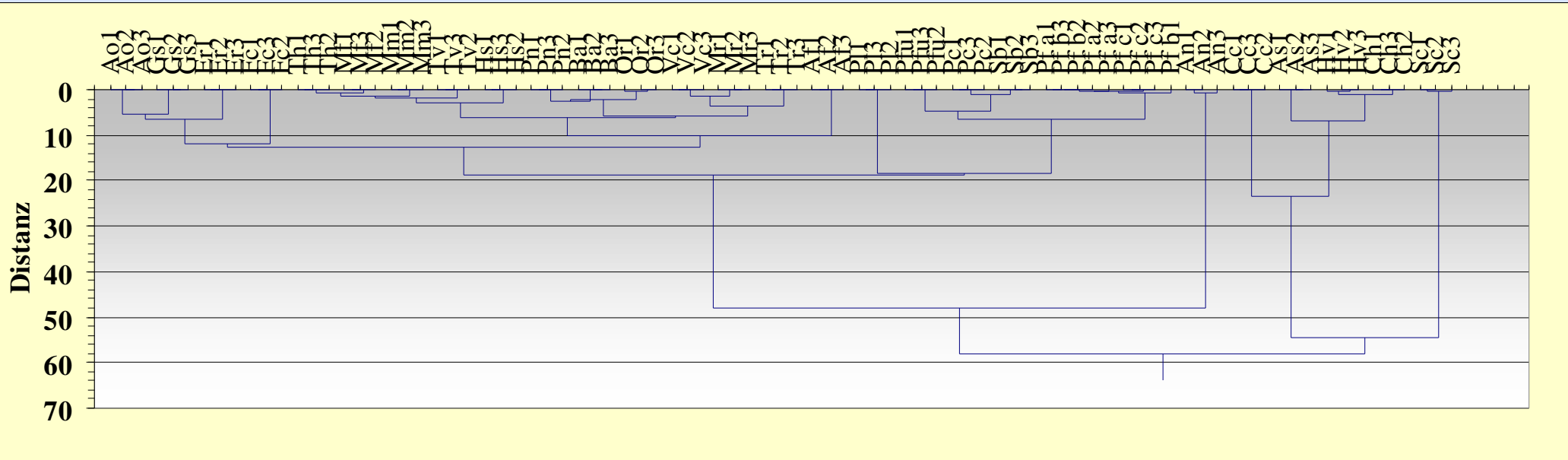
# Komplette Verbindung (Complete Linkage)



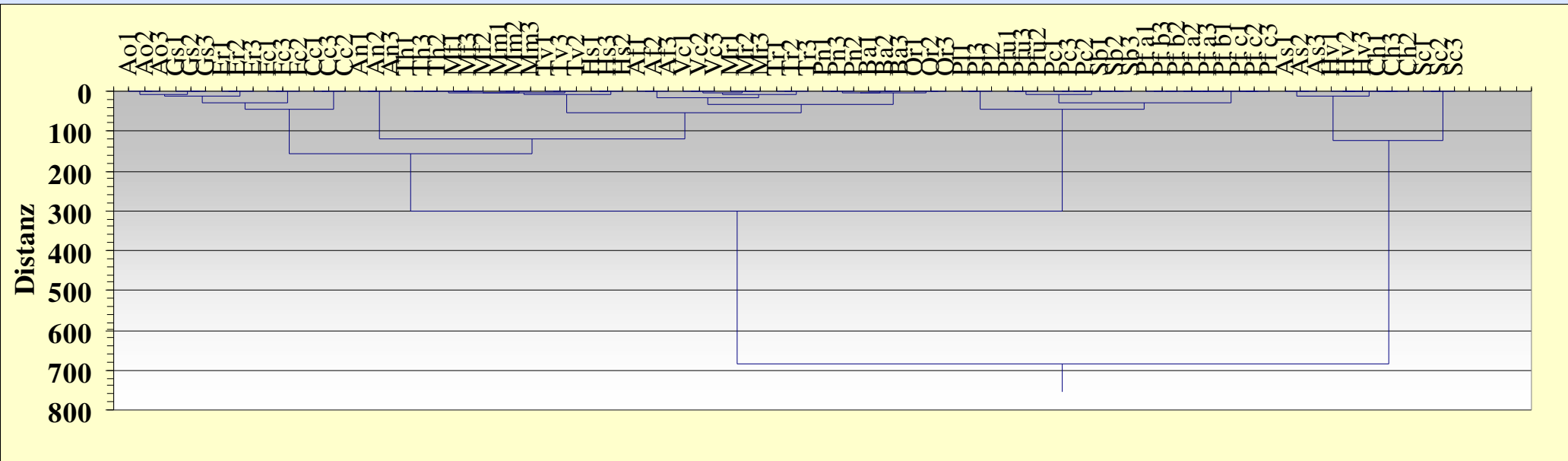
# Durchschnittsverbindung (Average Linkage)



# Zentroidverbindung (Centroid Linkage)

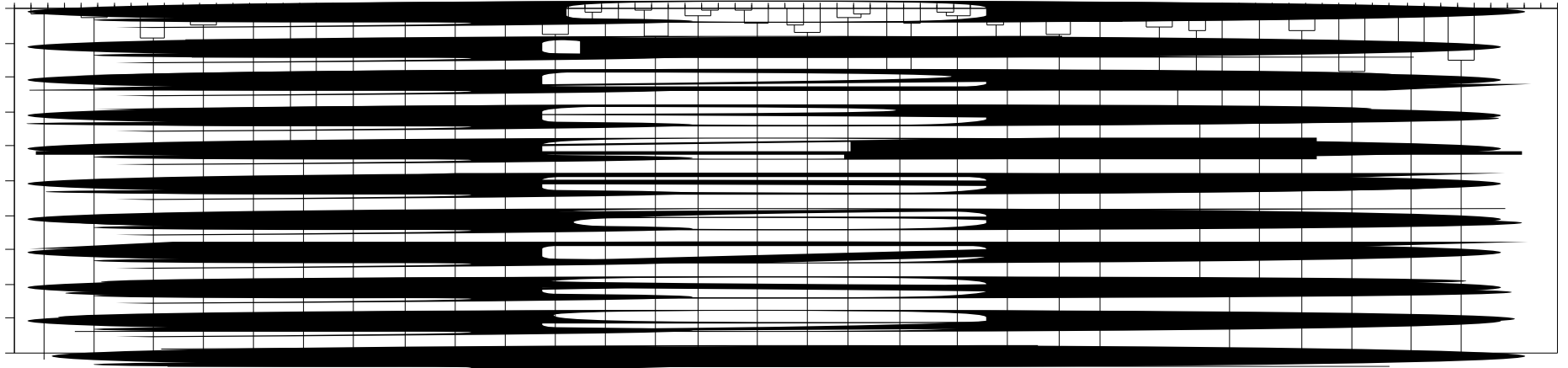


# Ward-Methode (Incremental Sums of Squares)



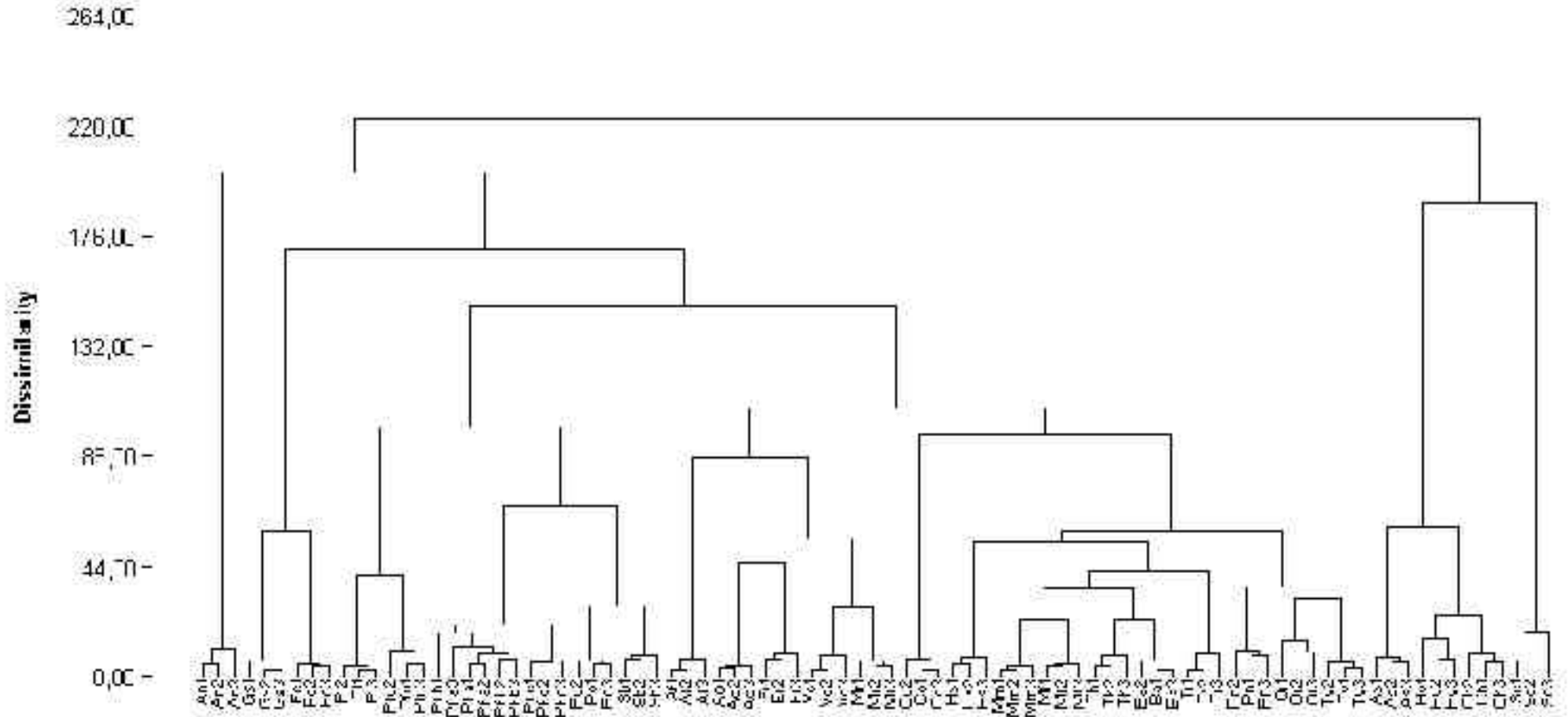
# Dendrogramm geclusterter Bilddaten von 29 Pilzstämmen (Ward)

*P. fellutanum*

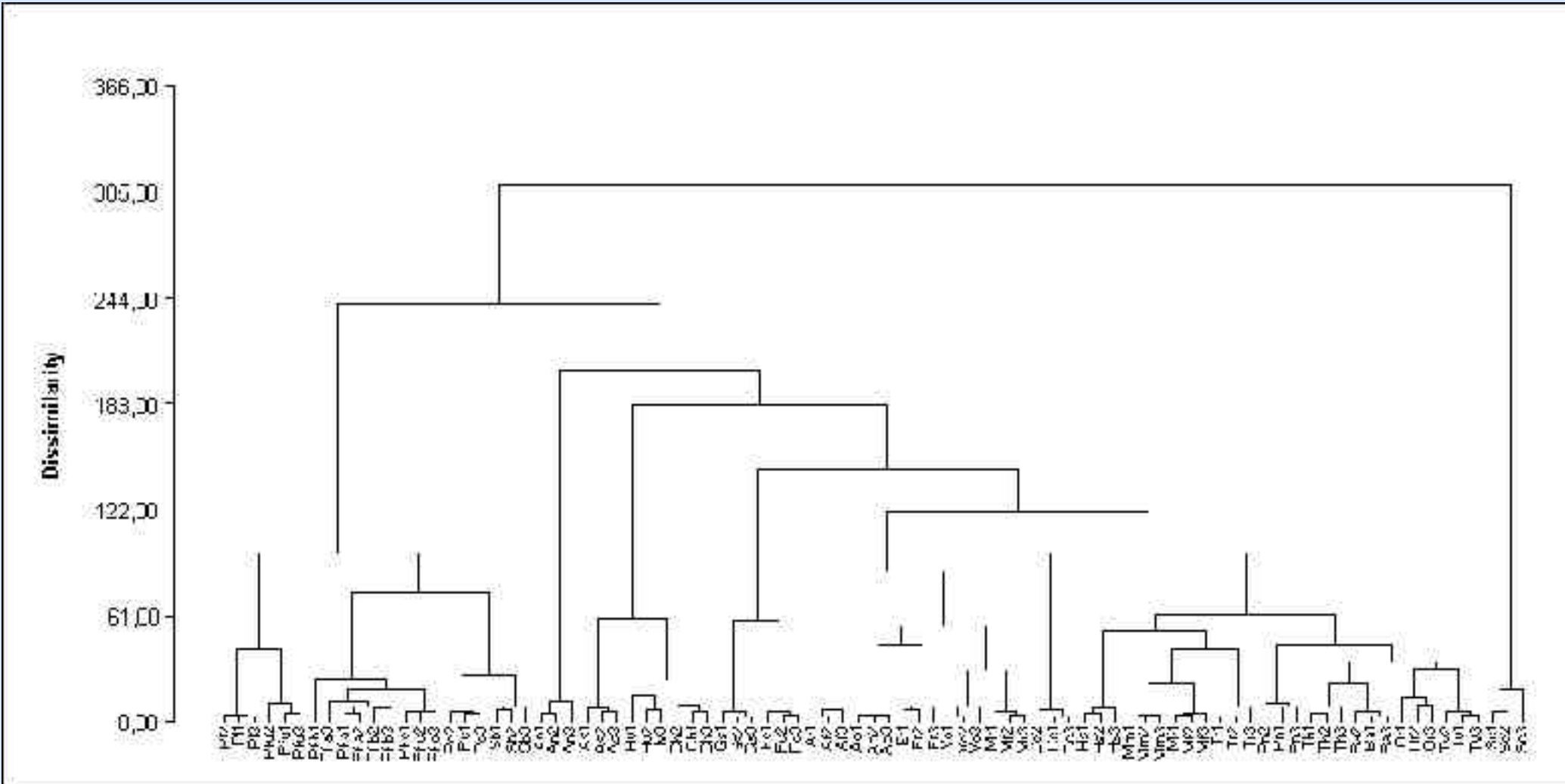


*Alternaria* sp., As; *Aspergillus flavus*, Af; *Aspergillus niger*, An; *Aspergillus ochraceus*, Ao; *Botrytis allii*, Ba; *Cladosporium cucumerinum*, Cc; *Cladosporium herbarum*, Ch; *Eurotium repens*, Er; *Fusarium culmorum*, Fc; *Gliocladium* sp., Gs; *Hormodendrum violaceum*, Hv; *Hypomyces* sp., Hs; *Mortierella ramanniana*, Mr; *Mucor flavus*, Mf; *Mucor mucedo*, Mm; *Oidiodendron rhodogenum*, Or; *Paecilomyces lilacinus*, Pl; *Penicillium camemberti*, Pc; *Penicillium funiculosum*, Pfu; *Penicillium fellutanum* (Isolat 1), Pf a; *Penicillium fellutanum* (Isolat 2), Pf b; *Penicillium fellutanum* (Isolat 3), Pf c; *Penicillium notatum* (Fleming Stamm), Pn; *Scopulariopsis brevicaulis*, Sb; *Stachybotrys chartarum*, Sc; *Trichoderma harzianum*, Th; *Trichoderma viride*, Tv; *Trichothecium roseum*, Tr; *Verticillium cinnabarinum*, Vc

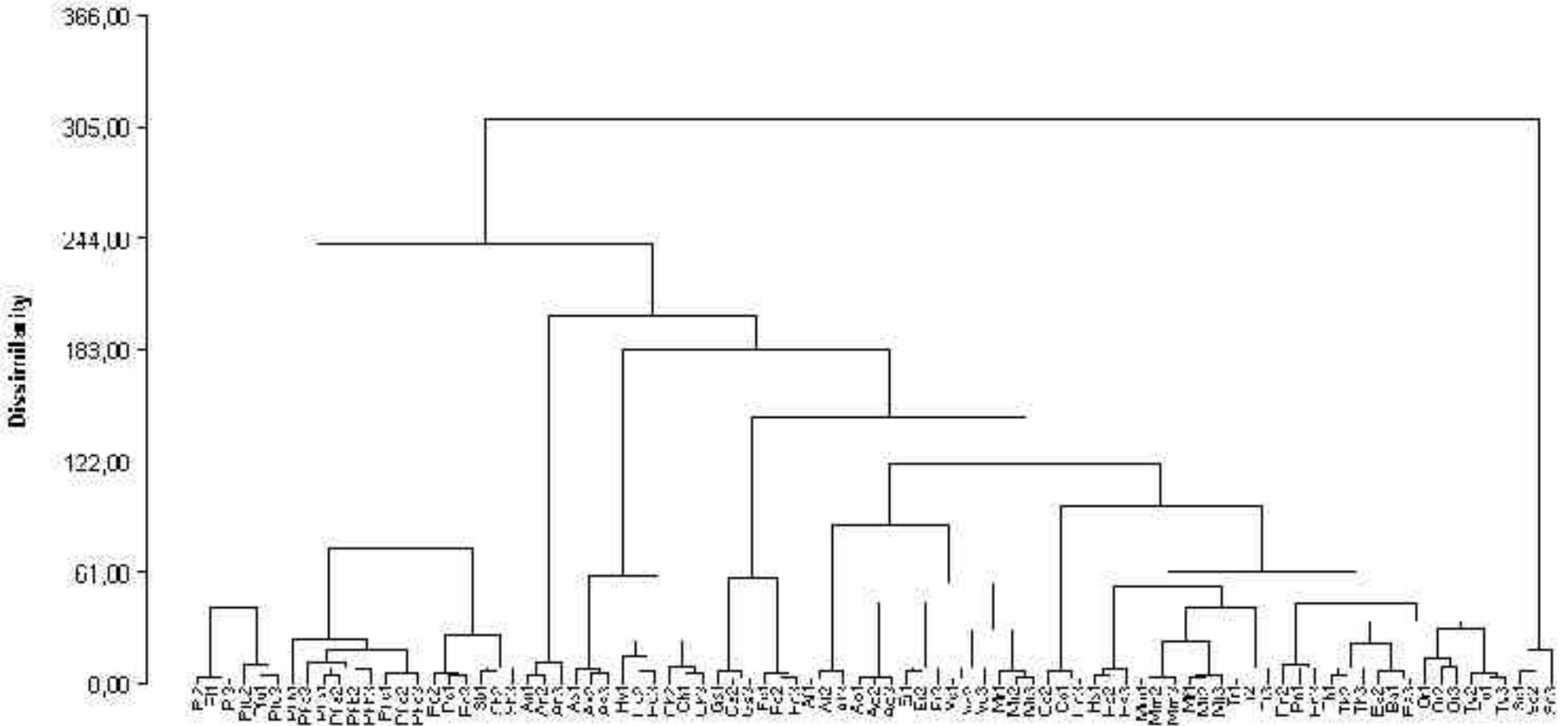
# Unweighted Pair Group Method with Arithmetic Mean



# Weighted Pair Group Method with Arithmetic Mean

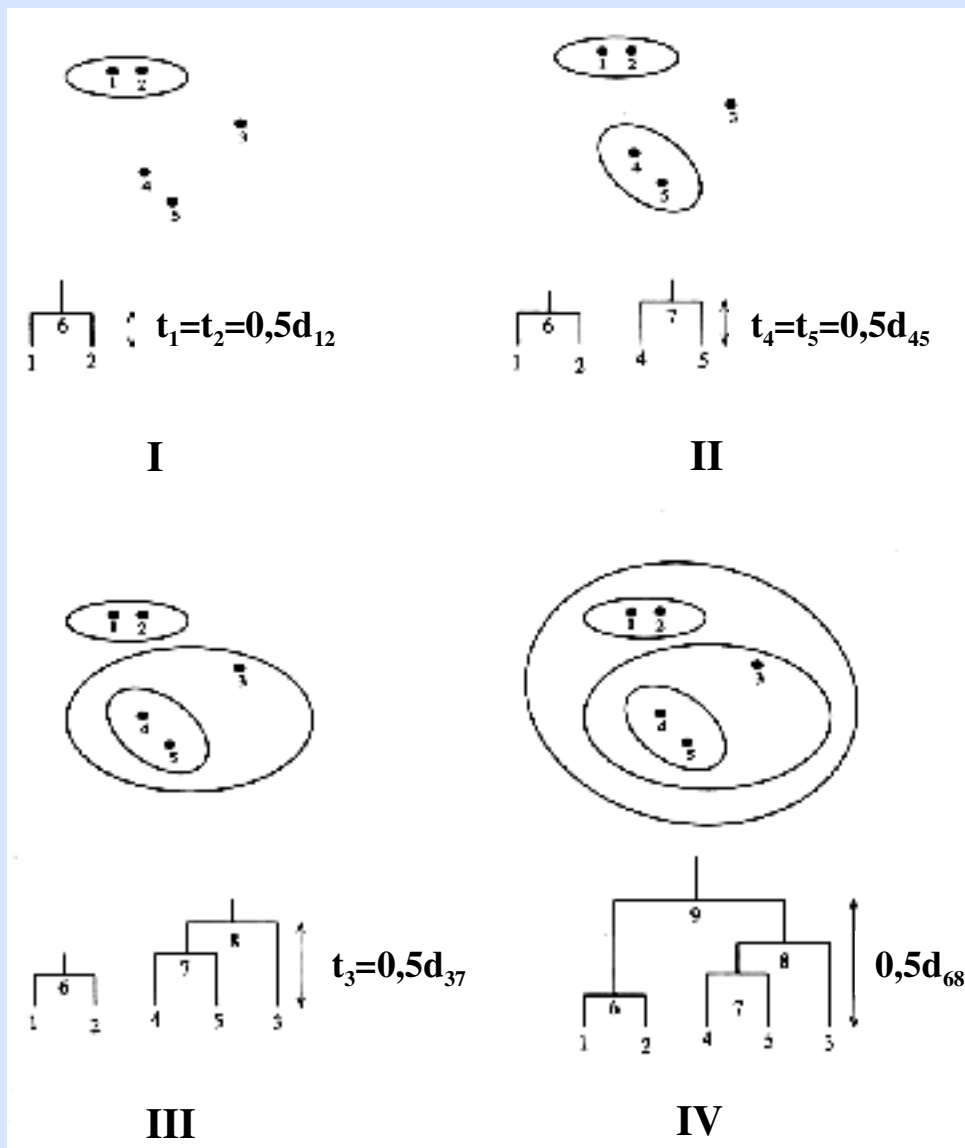


# Flexible Linkage





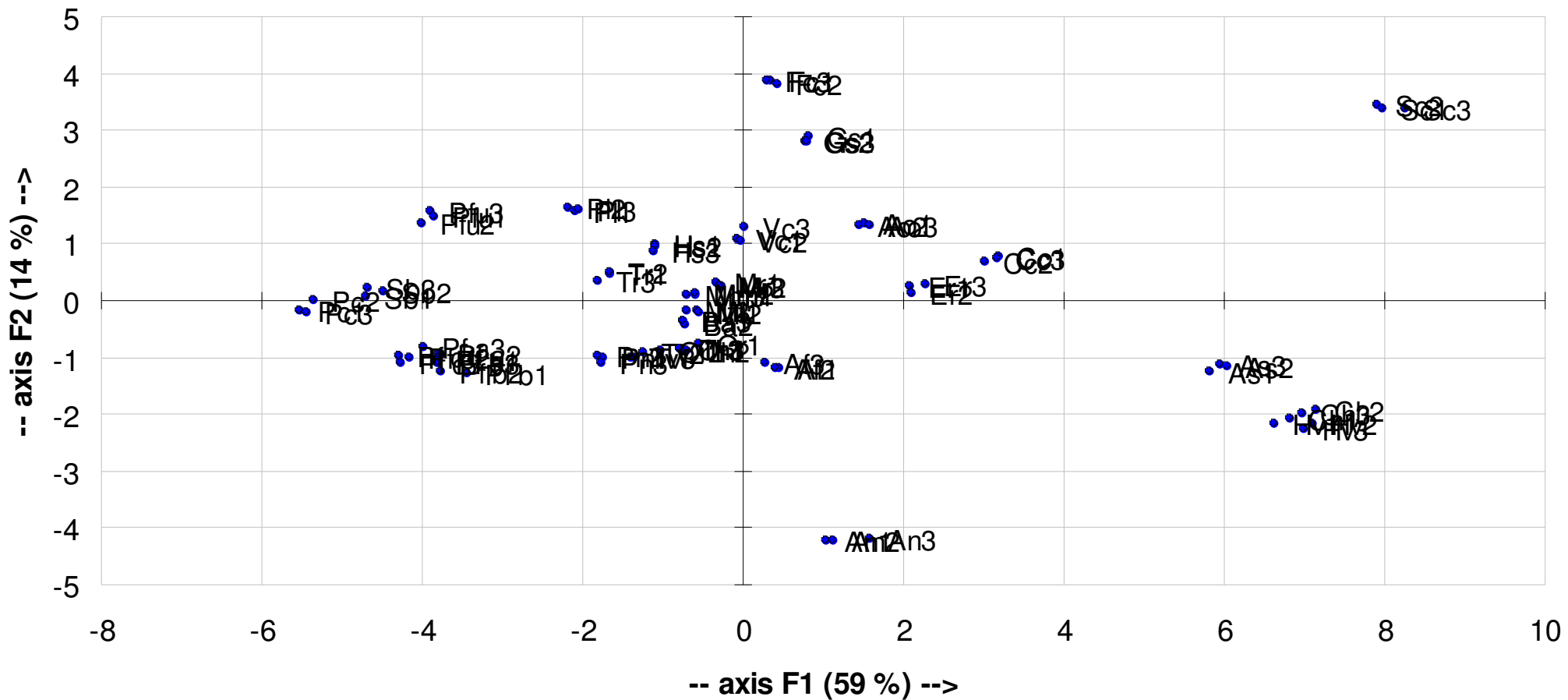
# Beispiel für die Erstellung eines *rooted trees* durch aufeinanderfolgende Clustering von Sequenzen (UPGMA)



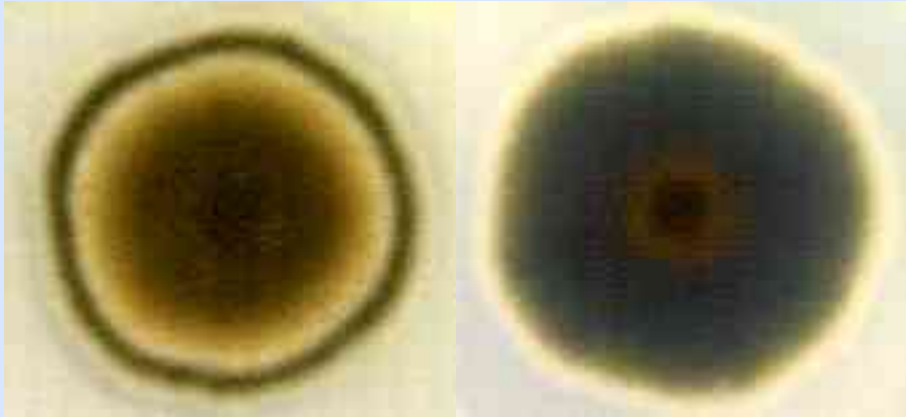
**d = distance**  
**t = edge lengths**

# Hauptkomponentenanalyse (Principal Components Analysis)

Observations (axis F1 and F2: 73 %)



# Ober- und Unterseite von Pilzkolonien unterschiedlicher Gattung



*Alternaria* sp.



*Gliocladium* sp.



*Paecilomyces lilacinus*



*Penicillium fellutanum*

# Makroskopische Bestimmung der Diversität von Pilzen

---

## **Drei Möglichkeiten der Einordnung:**

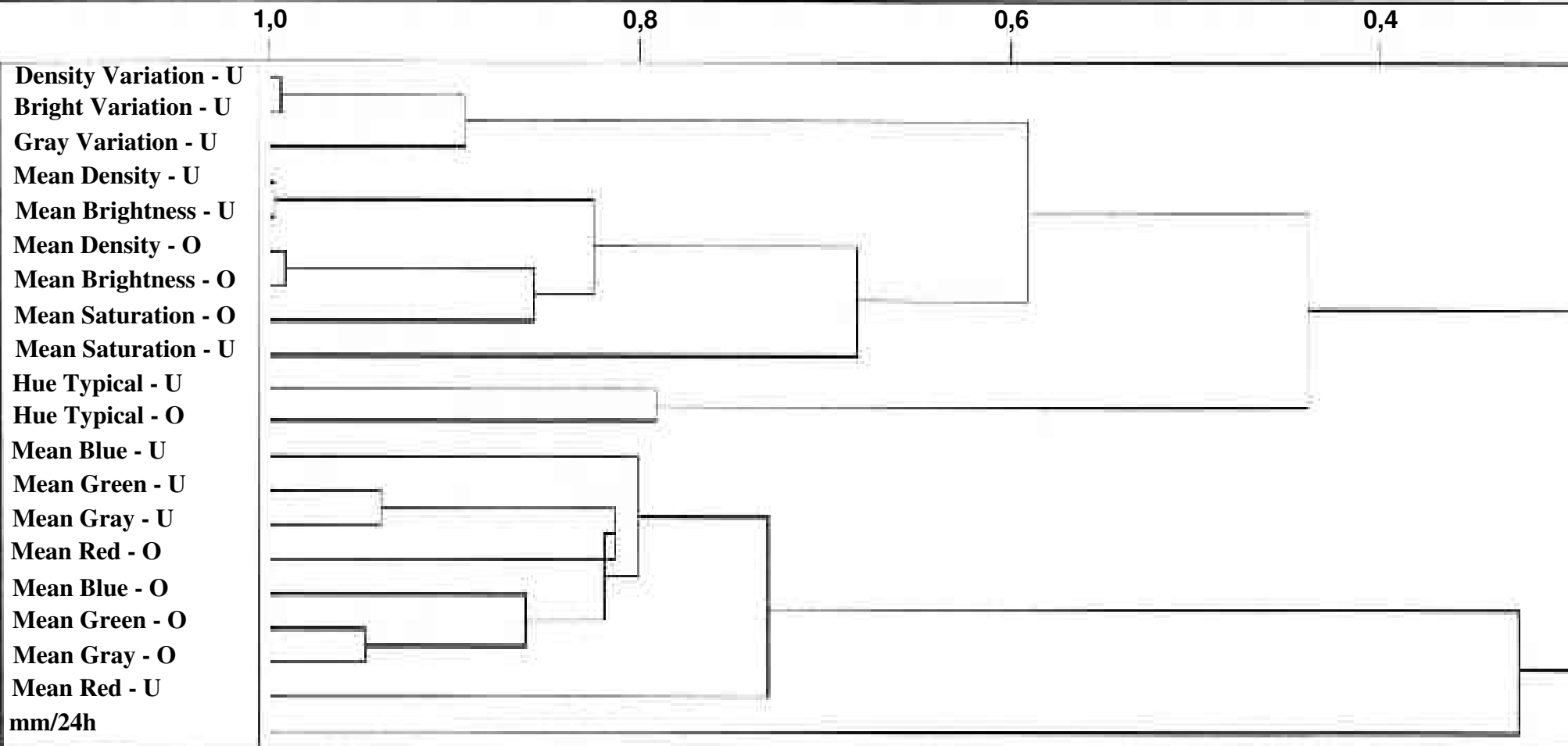
- a) subjektive Betrachtung lebender Kulturen (auf Röhrrchen/Platten)
- b) subjektive Betrachtung von Fotografien oder digitalen Bildern der Kulturen
- c) objektive Auswertung von Bilddaten (Werte für Wachstum und Farben) durch Hauptkomponenten- oder Clusteranalyse mit Ausgabe 2- oder 3- dimensionaler Diagramme bzw. von Dendrogrammen

## Voraussetzungen für die Clusteranalyse von Bilddaten:

- simultane Analyse zu vergleichender Datensätze (z.B. aus zwei Böden)
- Festlegung der optimalen Clusterebene:  
nachträgliche Zusammenführung kleiner Cluster sinnvoller als Trennung großer Cluster

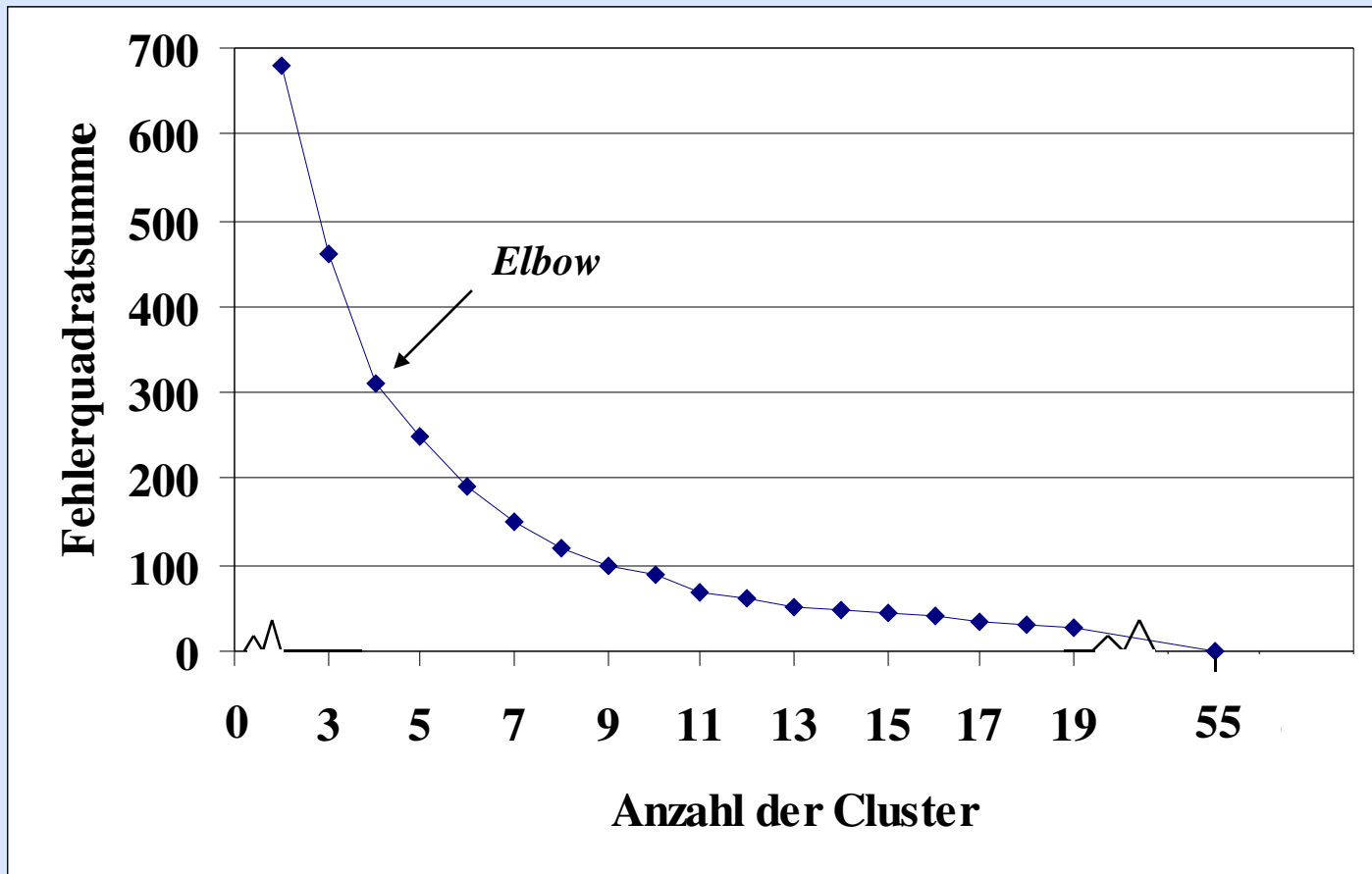
# Vergleich der Unterscheidungskraft verwendeter Meßgrößen

## Ähnlichkeit

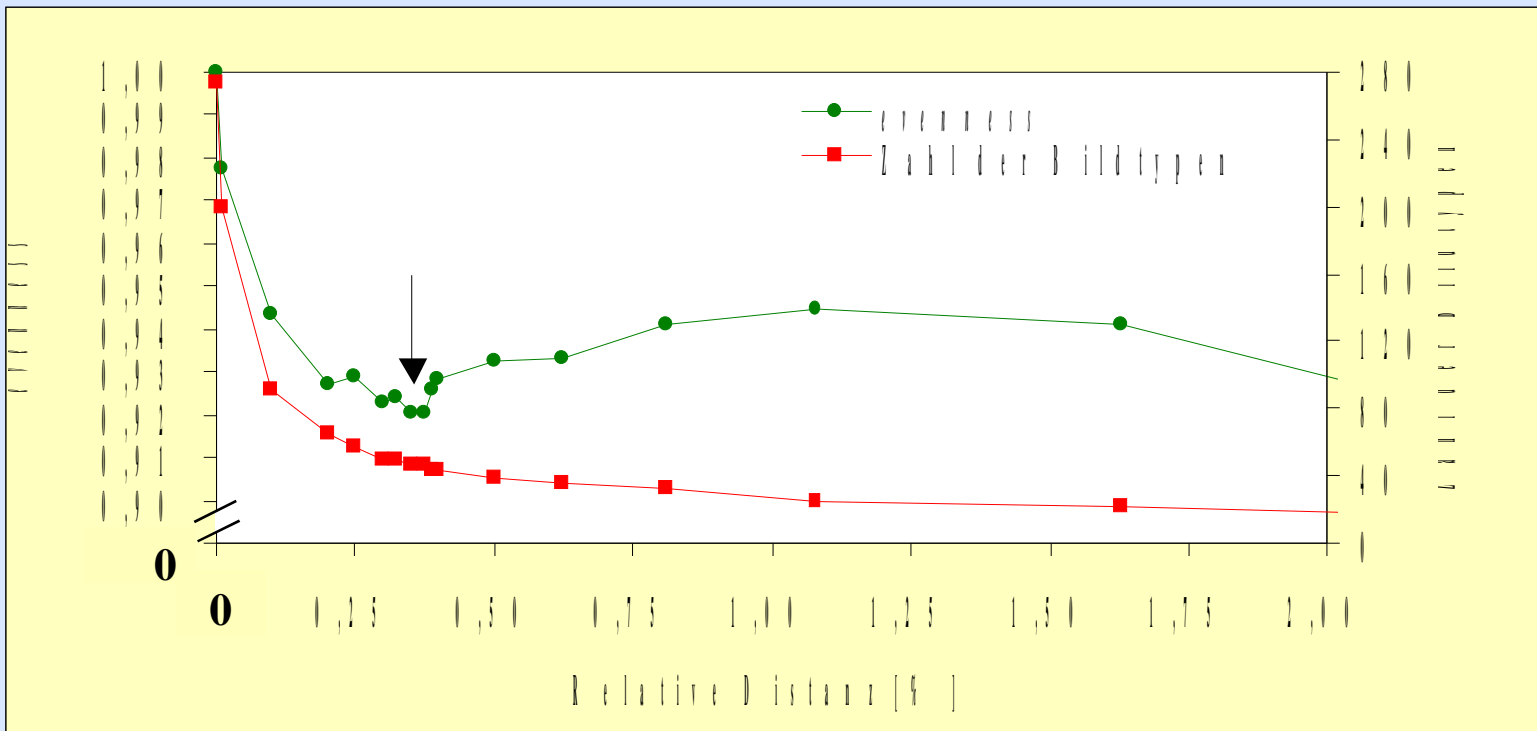


# Elbow-Kriterium zur Bestimmung der Clusteranzahl

Heterogenitätsmaß: Varianzkriterium = Fehlerquadratsumme



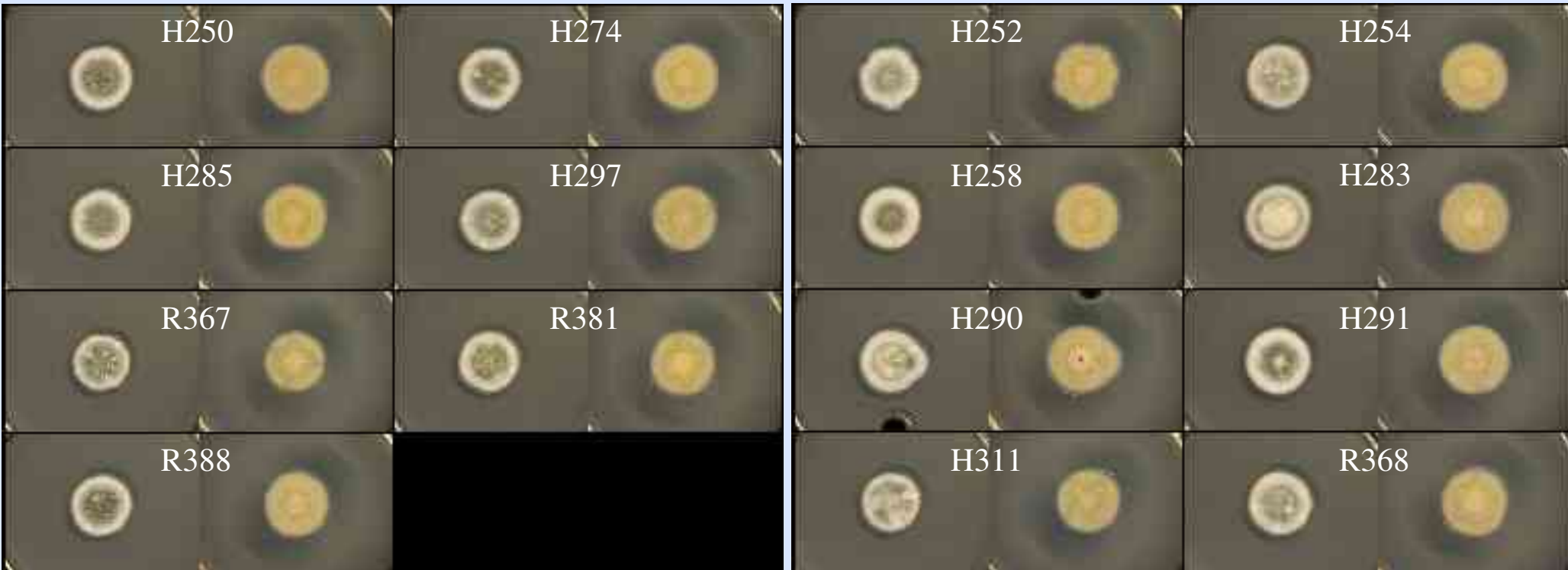
# Ermittlung der geeigneten Clusterebene: relative Distanz [%] in Dendrogrammen und *evenness* von Bildtypen in Stichproben



# Überprüfung der korrekten Gruppierung von Bilddaten anhand der zugehörigen Aufnahmen I

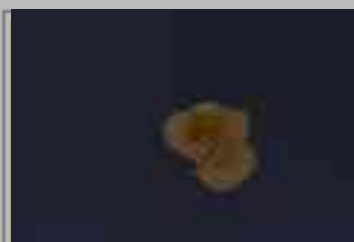
Cluster I

Cluster II

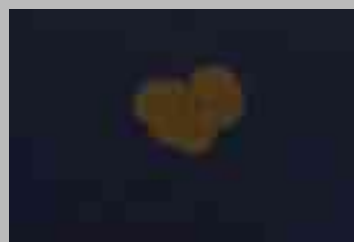


# Überprüfung der korrekten Gruppierung von Bilddaten anhand der zugehörigen Aufnahmen II

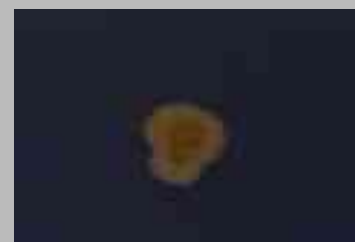




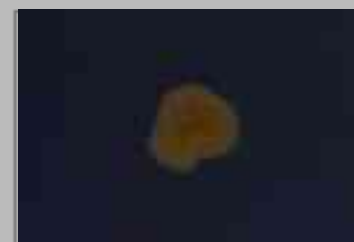
nr5hn lim



nr5h u lim



nr5cn lim



nr5c u lim



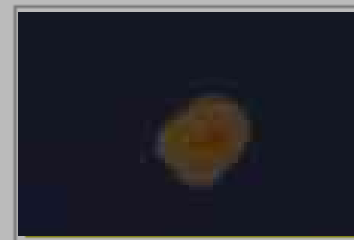
nr5dn lim



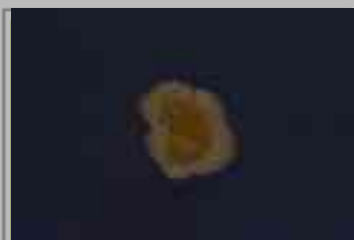
nr5d u lim



nr5en lim



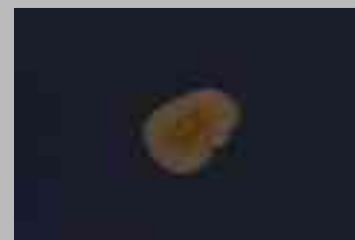
nr5e u lim



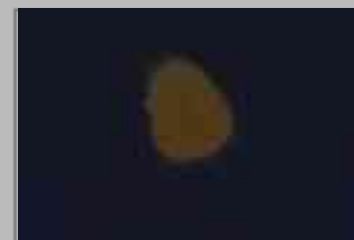
nr5fn lim



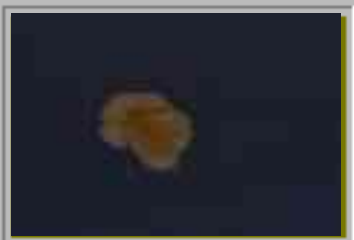
nr5fu lim



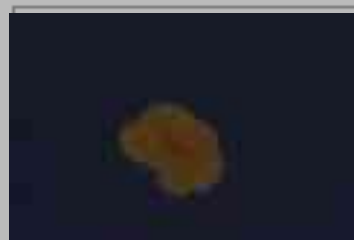
nr5gn lim



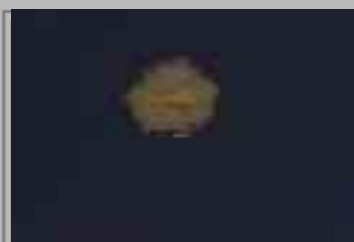
nr5g u lim



nr5n lim



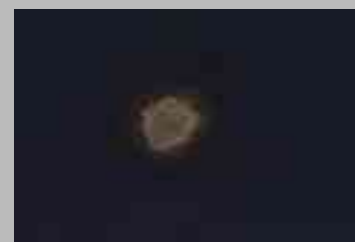
nr5u lim



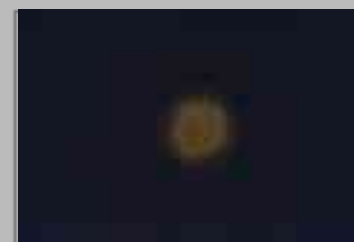
h100n lin



h170u lin



F34n lin



h34u lin



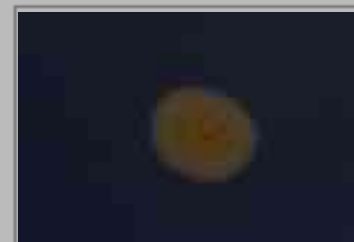
F75n lin



h75u lin



F78n lin



h78u lin

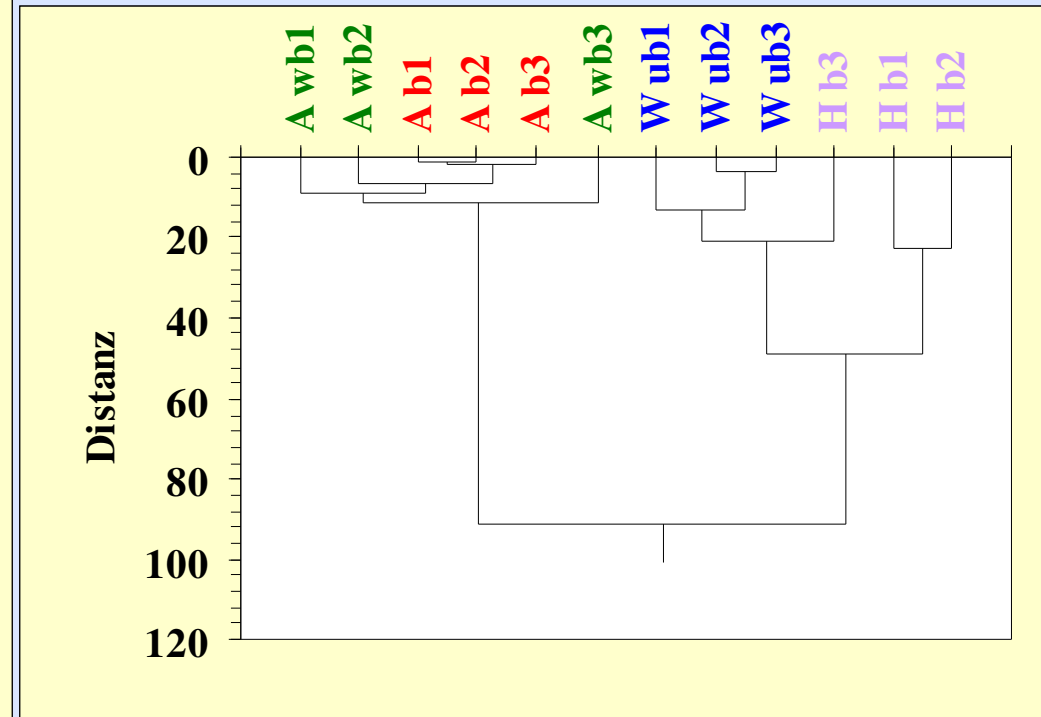
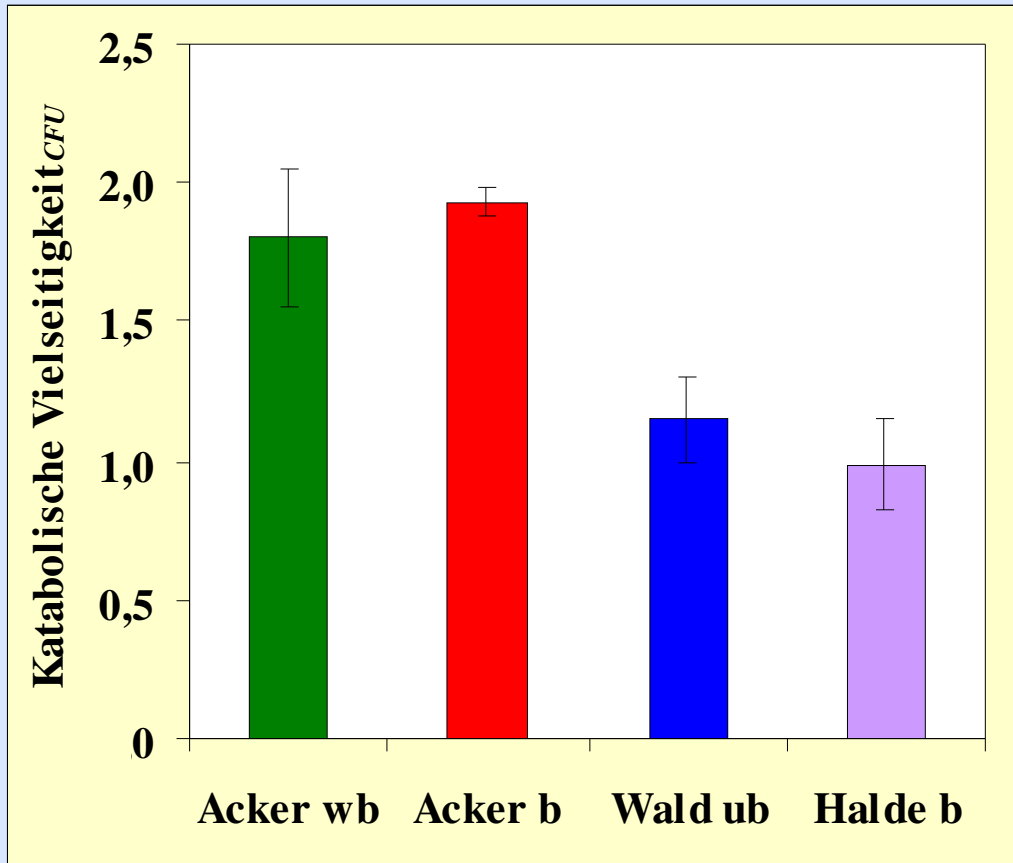


F89n lin



h89u lin

# Abschätzung der katabolischen Vielseitigkeit von Bodenpilzen aus der Verteilung aromatspezifischer Vermehrungseinheiten (*CFU*)

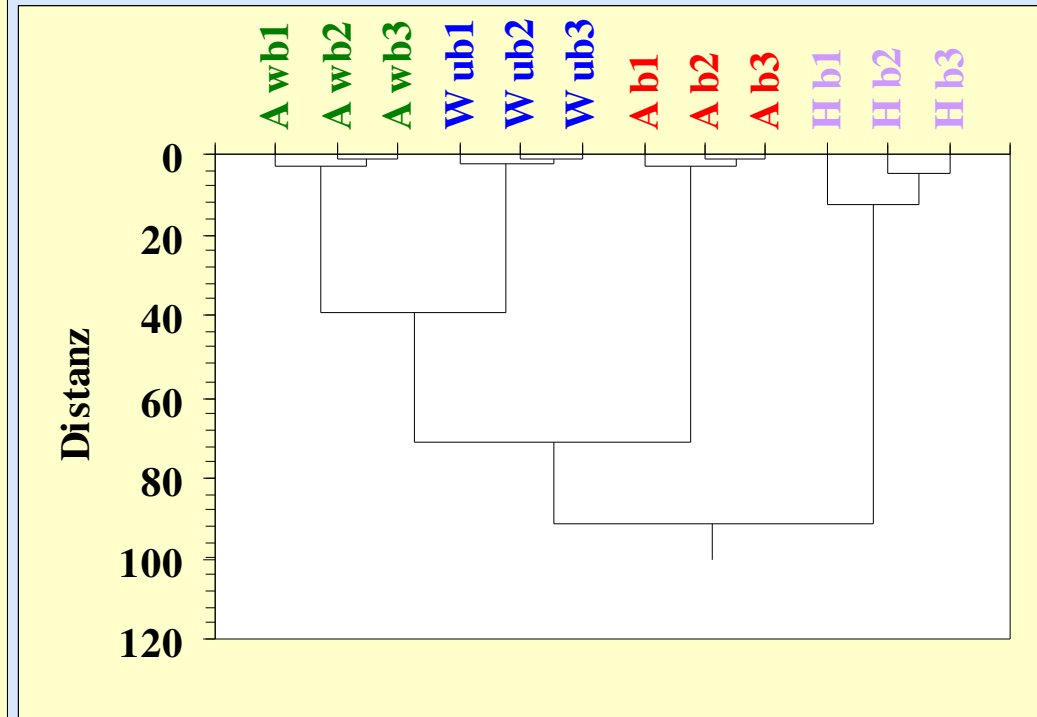
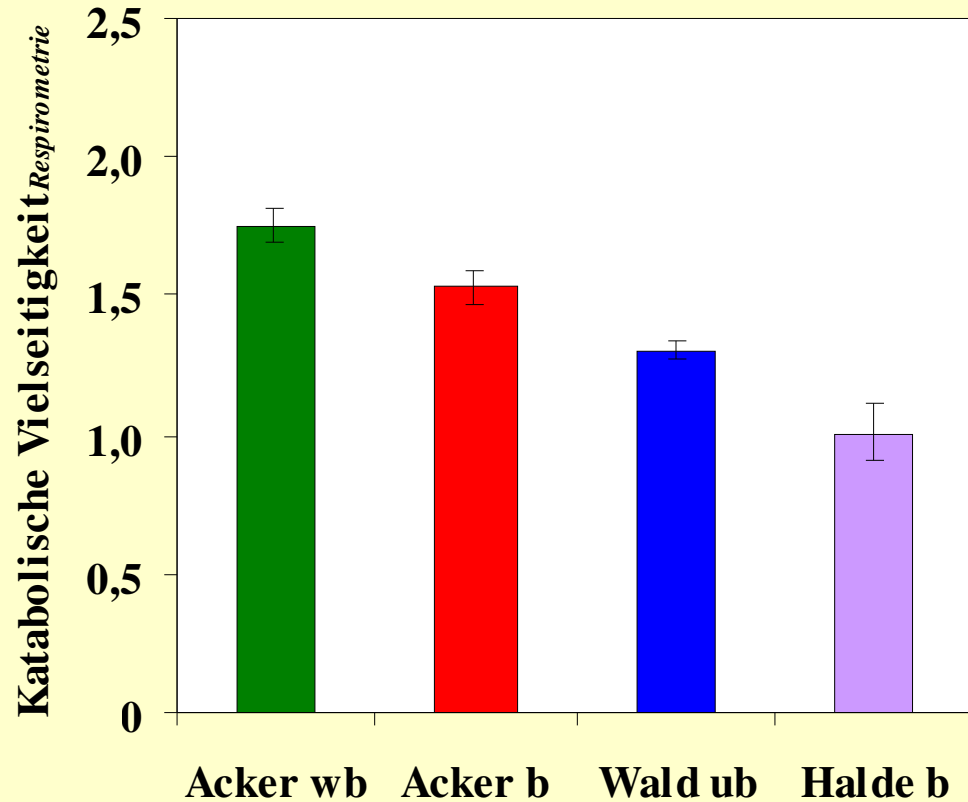


Bodenpaar I: Ackerböden (A), wb = wenig belastet, b = belastet.

Bodenpaar II: Wald- (W) und Haldenböden (H), ub = unbelastet, b = belastet.

*CFU* = colony forming units.

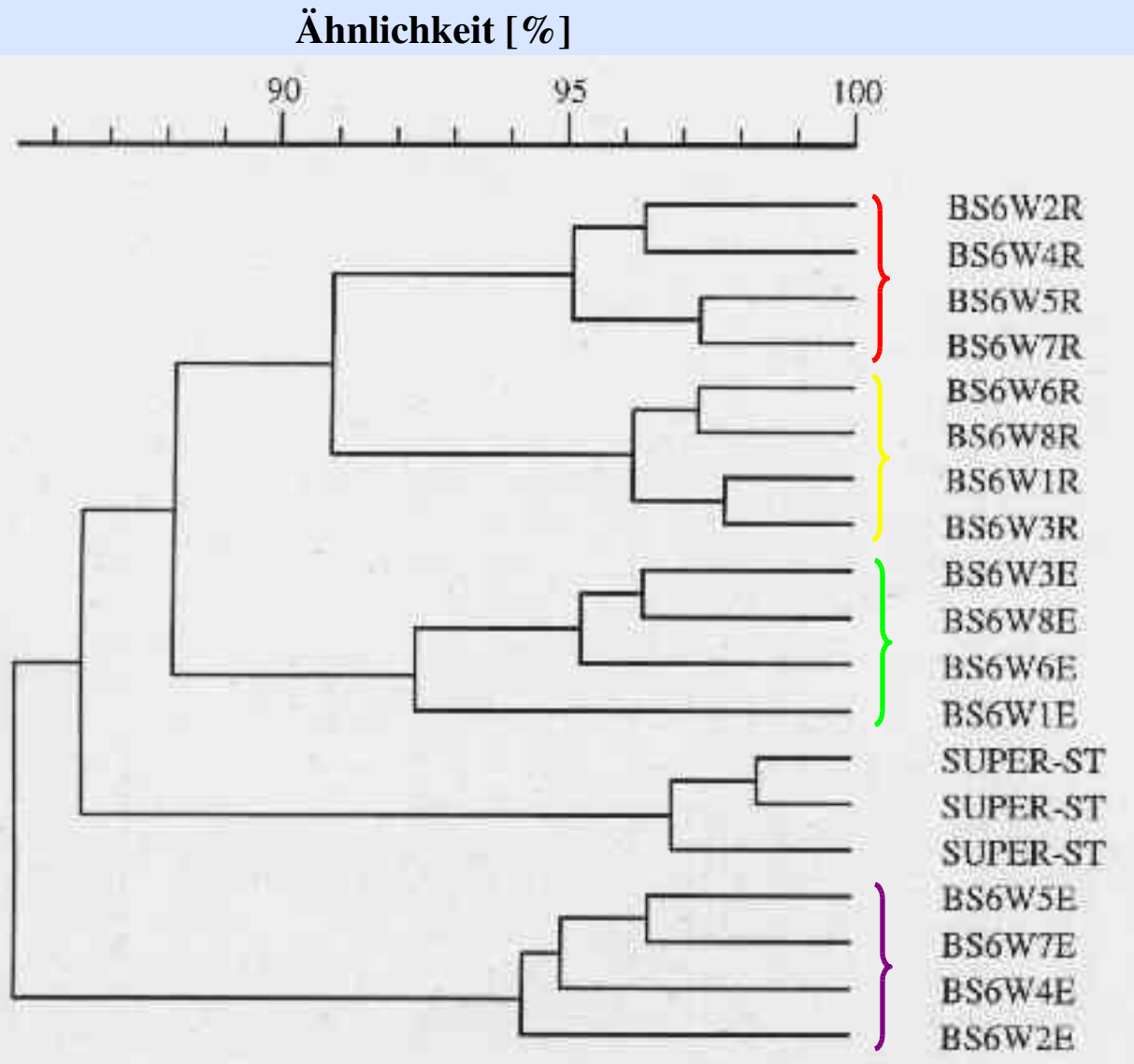
# Abschätzung der katabolischen Vielseitigkeit von Böden durch Verrechnung der reziproken Abbauzeiten aromatischer Säuren



Bodenpaar I: Ackerböden (A), wb = wenig belastet, b = belastet.

Bodenpaar II: Wald- (W) und Haldenboden (H), ub = unbelastet, b = belastet.

# Vergleich der Pilzvielfalt von Rhizosphären- und Endophyten-Gemeinschaften bei Raps und Erdbeere am Standort BS (DGGE)



## Ward-Methode

**DGGE:** Denaturierende Gradienten  
Gel Elektrophorese

**Super-St.:** Banden-Standard,  
bestehend aus 11  
Pilzisolaten

**BS6:** 6. Probenahme BS

**W1,3,6,8:** Wurzelproben Raps

**W2,4,5,7:** Wurzelproben Erdbeere

**E:** Endophytische Pilze

**R:** Rhizosphärische Pilze

# Zusammenfassung

## Hierarchische Verfahren

- Start feinste Partitionierung, jedes Objekt = 1 Cluster
- Clusterbildung Fusionierung von Clustern
- Ziel Kriterium erfüllt
- Im Dendrogramm gesamte Gruppierungsinformation enthalten  
Problem: Findung geeigneter Clusterebene
- Ward-Methode als Einstieg: sehr gut

## Nicht-hierarchische Verfahren

(Partitionierendes Cluster-Verfahren, Clusterzentrenanalyse)

Vorgabe Startgruppierung

Verschieben/Sortieren der Objekte

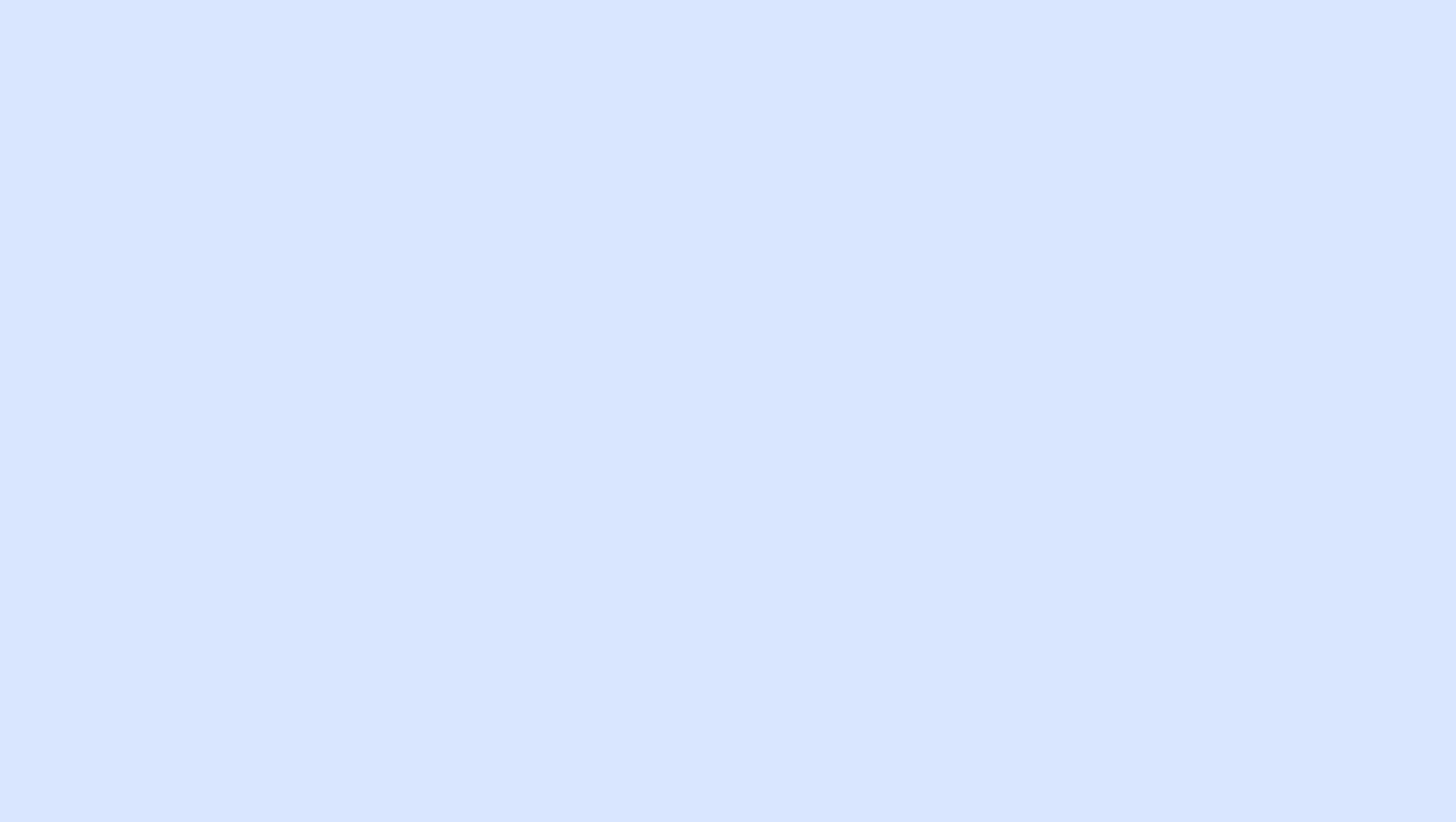
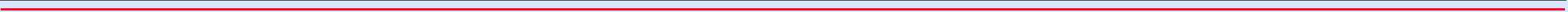
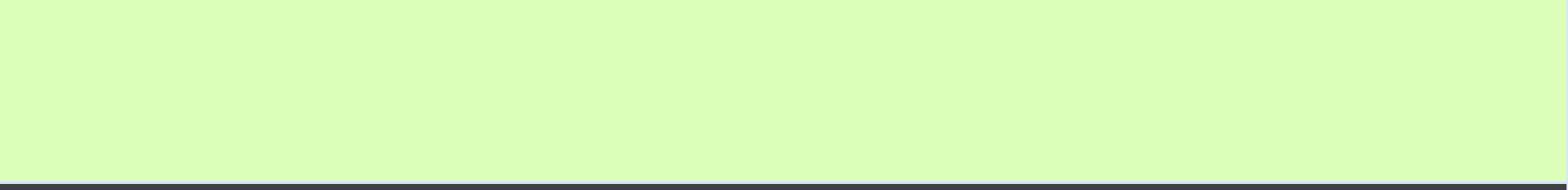
Kriterium erfüllt

- Problem: enormer Arbeits- + Zeitaufwand, daher meist nur Annäherungen möglich → bei 10 Objekten 115 975 verschiedene Möglichkeiten

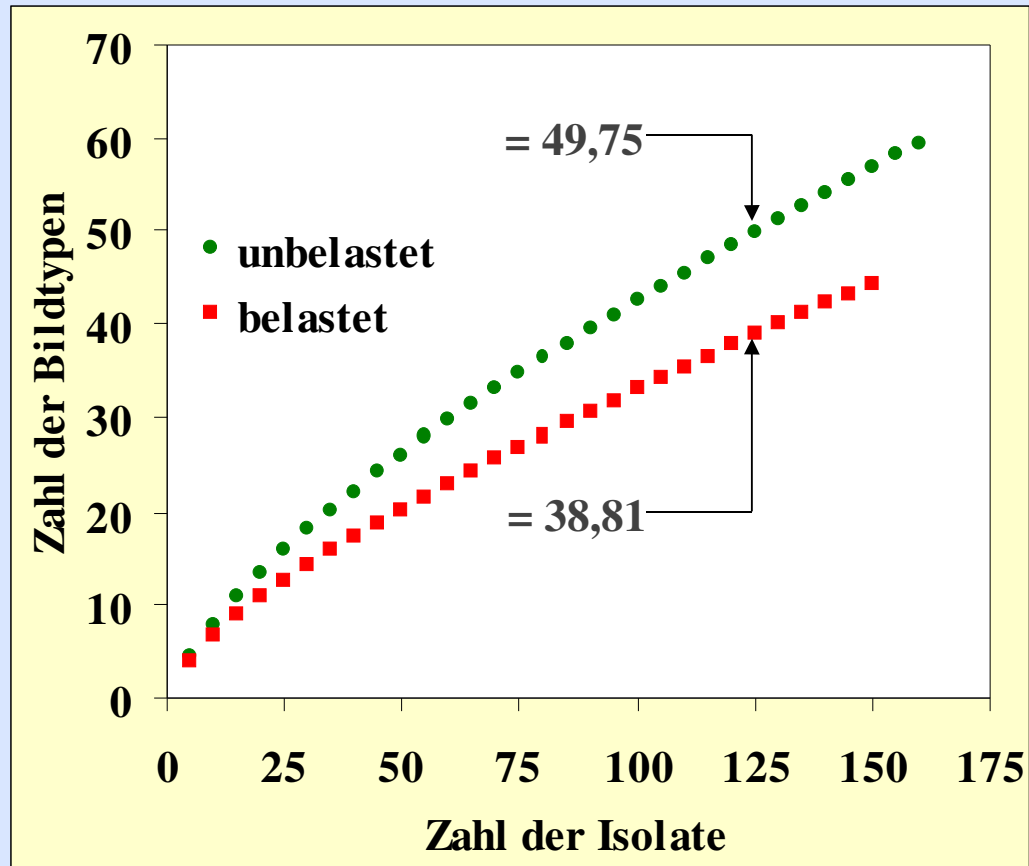
-Testung verschiedener Verfahren empfohlen →

Gefahr der *geschönten* Darstellung von Daten

bzw. Verzerrung



# Korrektur Vergleich des Artenreichtums von Stichproben durch *rarefaction*



# Ökologische Indices

**Artenreichtum:**  $d = \frac{S}{N}$

**Diversität:**  $\bar{H} = -\sum \left( \frac{n_i}{N} \right) \times \log \left( \frac{n_i}{N} \right)$  **Shannon-Index** (Shannon und Weaver, 1963)

= Vielfalt; Artenzahl in Lebensgemeinschaft,

wobei S = Zahl der Arten (Bildtypen)

N = Zahl der Individuen (Isolate)

$n_i$  = Zahl der Individuen je Art bzw. Bildtyp,  $\frac{n_i}{N}$  = Bedeutungswert der Art.

$$e = \frac{\bar{H}}{\log A}$$

**evenness:**

(Pielou, 1966)

$\bar{H}$  ≡ Gleichmäßigkeit der Verteilung (z.B. von Arten, katabol. Fähigkeiten),

wobei  $\bar{H}$  = Shannon-Index A = Zahl der angebotenen Substrate

# Beispielrechnung Diversität und *evenness*

| Modelle gleicher Arten- und Individuenzahl, aber unterschiedlicher Artenverteilung: Berechnung von <b>Diversität</b> und <i>evenness</i> |                 |           |                            |                        |
|--|-----------------|-----------|----------------------------|------------------------|
| Individuenz  |                 | Bedeutung | -log Bedeutung             | Produkt                |
| $n_i$  | $n_i$           | $n_i/N$   | $-\log(n_i/N)$             | $-(n_i/N) \log(n_i/N)$ |
| 3  | 1               | 0,0048    | 2,3222                     | 0,0111                 |
| 4  | 1               | 0,0048    | 2,3222                     | 0,0111                 |
| 4  | 1               | 0,0048    | 2,3222                     | 0,0111                 |
| 5  | 1               | 0,0048    | 2,3222                     | 0,0111                 |
| 5  | 2               | 0,0095    | 2,0212                     | 0,0192                 |
| 6  | 2               | 0,0095    | 2,0212                     | 0,0192                 |
| 6  | 3               | 0,0143    | 1,8451                     | 0,0264                 |
| 7  | 3               | 0,0143    | 1,8451                     | 0,0264                 |
| 8  | 4               | 0,0190    | 1,7202                     | 0,0328                 |
| 9  | 5               | 0,0238    | 1,6232                     | 0,0386                 |
| 9  | 6               | 0,0286    | 1,5441                     | 0,0441                 |
| 10   | 8               | 0,0381    | 1,4191                     | 0,0541                 |
| 12   | 9               | 0,0429    | 1,3680                     | 0,0586                 |
| 13   | 12              | 0,0571    | 1,2430                     | 0,0710                 |
| 14   | 14              | 0,0667    | 1,1761                     | 0,0784                 |
| 15   | 17              | 0,0810    | 1,0918                     | 0,0884                 |
| 17   | 22              | 0,1048    | 0,9798                     | 0,1026                 |
| 19   | 26              | 0,1238    | 0,9072                     | 0,1123                 |
| 21   | 33              | 0,1571    | 0,8037                     | 0,1263                 |
| 23   | 40              | 0,1905    | 0,7202                     | 0,1372                 |
| <b>Summe N</b>   | <b>210</b>      |           | <b>Diversität (Summe)</b>  | <b>1,0799</b>          |
| <b>n = 20</b>  |                 |           | <b>log Artenzahl (20)</b>  | <b>1,3010</b>          |
| <b>e = 0,95</b>  | <b>e = 0,83</b> |           | <b>evenness (D/log AZ)</b> | <b>0,8301</b>          |

| Abbauzeit [h]  | Bedeutung    | -log Bedeutung             | Produkt                |
|----------------|--------------|----------------------------|------------------------|
| $f_i$          | $f_i/F$      | $-\log(f_i/F)$             | $-(f_i/F) \log(f_i/F)$ |
| 19             | 0,0244       | 1,6131                     | 0,0393                 |
| 22,3           | 0,0286       | 1,5436                     | 0,0442                 |
| 23             | 0,0295       | 1,5301                     | 0,0451                 |
| 51,5           | 0,0661       | 1,1801                     | 0,0780                 |
| 21,8           | 0,0280       | 1,5534                     | 0,0434                 |
| 59,5           | 0,0763       | 1,1174                     | 0,0853                 |
| 44             | 0,0564       | 1,2484                     | 0,0705                 |
| 30,3           | 0,0389       | 1,4104                     | 0,0548                 |
| 31,5           | 0,0404       | 1,3936                     | 0,0563                 |
| 34,3           | 0,0440       | 1,3566                     | 0,0597                 |
| 30,5           | 0,0391       | 1,4076                     | 0,0551                 |
| 64,3           | 0,0825       | 1,0837                     | 0,0894                 |
| 40,3           | 0,0517       | 1,2866                     | 0,0665                 |
| 48,3           | 0,0620       | 1,2079                     | 0,0748                 |
| 33             | 0,0423       | 1,3734                     | 0,0581                 |
| 32,3           | 0,0414       | 1,3827                     | 0,0573                 |
| 10,8           | 0,0139       | 1,8584                     | 0,0257                 |
| 41,3           | 0,0530       | 1,2759                     | 0,0676                 |
| 48,8           | 0,0626       | 1,2035                     | 0,0753                 |
| 59,8           | 0,0767       | 1,1152                     | 0,0855                 |
| 33             | 0,0423       | 1,3734                     | 0,0581                 |
| <b>Summe F</b> | <b>779,6</b> | <b>Diversität (Summe)</b>  | <b>1,2901</b>          |
| <b>n = 21</b>  |              | <b>log 21</b>              | <b>1,3222</b>          |
|                |              | <b>evenness (D/log 21)</b> | <b>0,9757</b>          |