Kristan Schneider

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

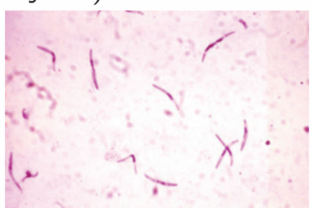
How often are you hit by an infection? - A likelihood approach to determine the multiplicity of infection

ROeS 2013, Sept. 12, 2013

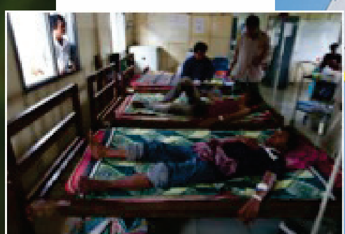
What is malaria?

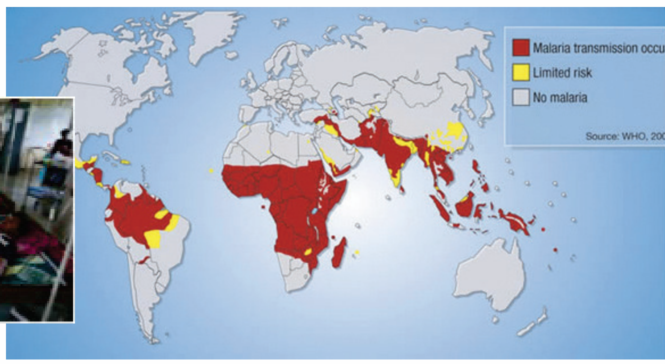
HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Infectious disease caused by **parasites** (genus ***Plasmodium***; eukaryote)

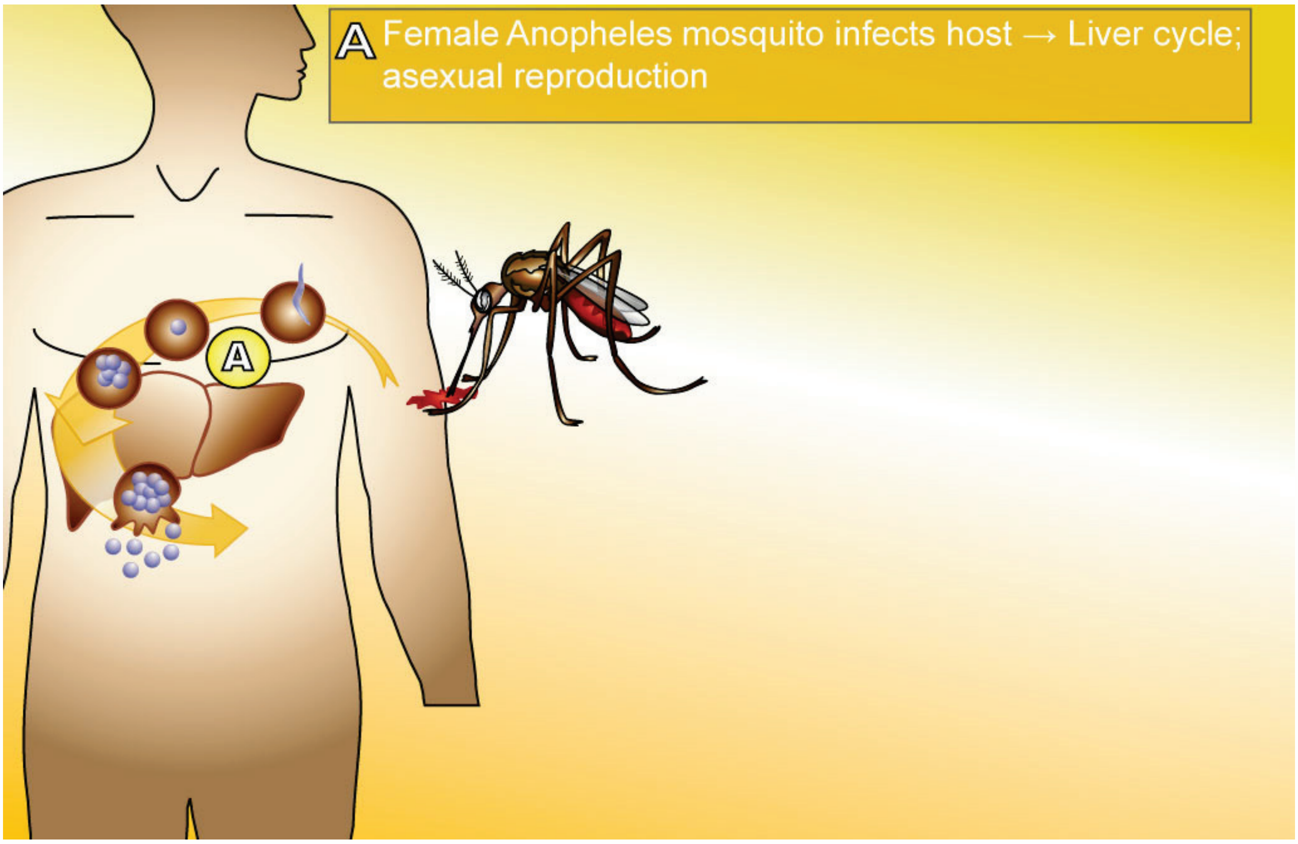


- 40% of the world's population at malaria risk
- Worldwide 200 - 300 million infections & 1-3 million deaths per year
- Enormous economic damage every year



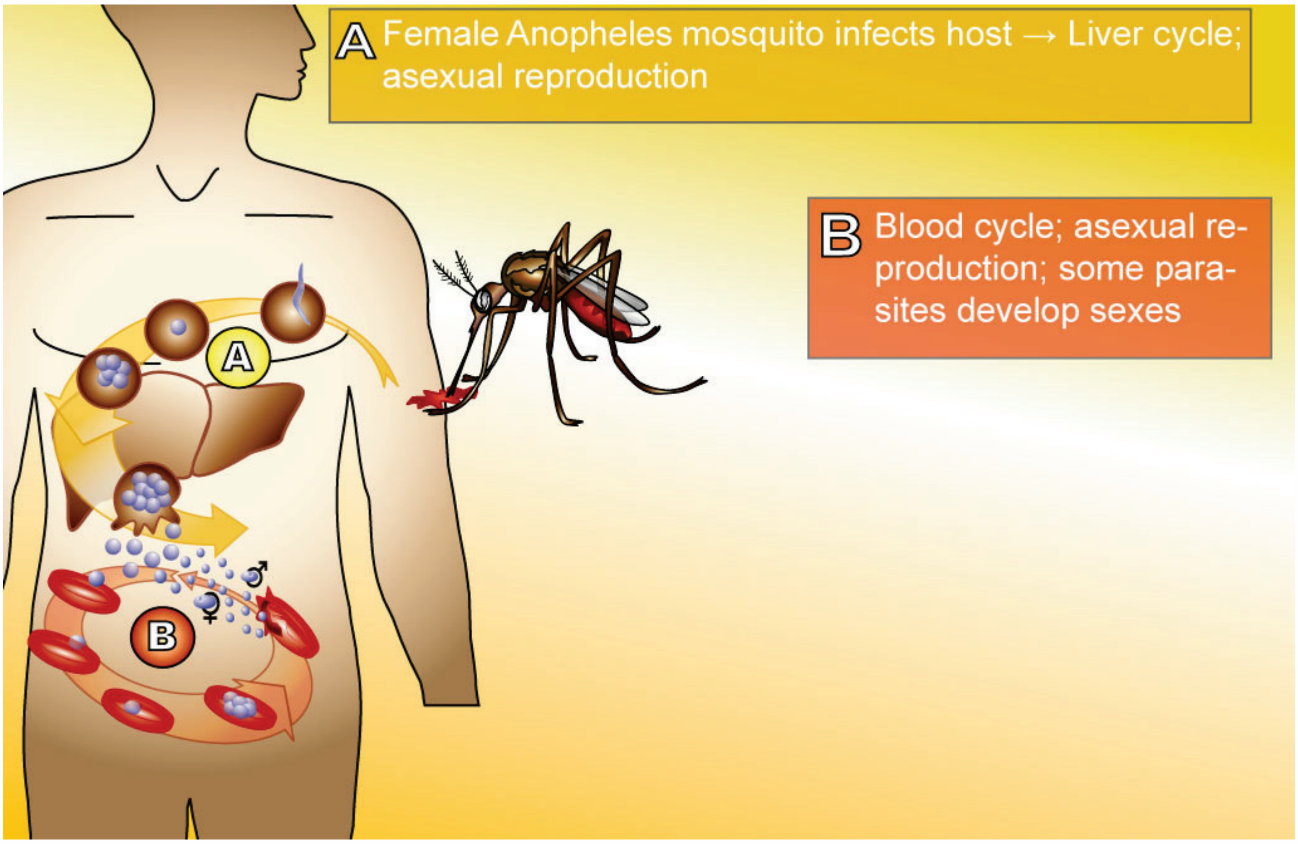


Transmission cycle

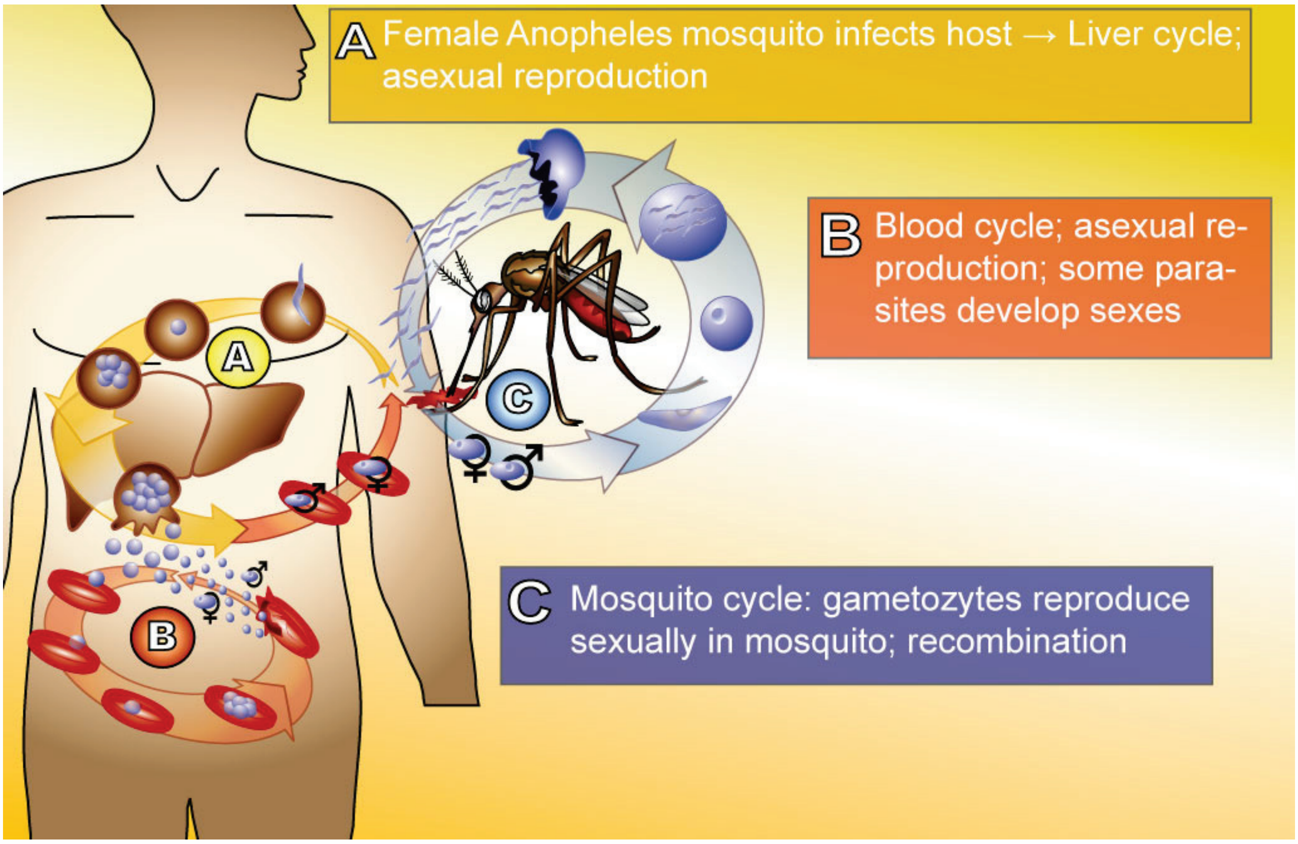


Transmission cycle

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES



Transmission cycle



Malaria control

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Malaria control

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

A: Multiplicity of infection = metric for transmission intensity

Malaria control

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

A: Multiplicity of infection = metric for transmission intensity

→ co-infections might increase disease severity

Malaria control

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

A: Multiplicity of infection = metric for transmission intensity

→ co-infections might increase disease severity

→ metric transmission intensity


Malaria control

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

A: Multiplicity of infection = metric for transmission intensity

→ co-infections might increase disease severity

→ metric transmission intensity

→ impact statistics based on genetic data


Malaria control

- Goal of **malaria control:**
  - reduce decease burden ⇒ drug treatments
  - reduce transmission ⇒ bed nets, vector control,...

Q: How to measure efficiency of control interventions?

A: Multiplicity of infection = metric for transmission intensity

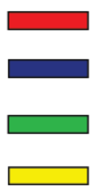
→ co-infections might increase disease severity

→ metric transmission intensity

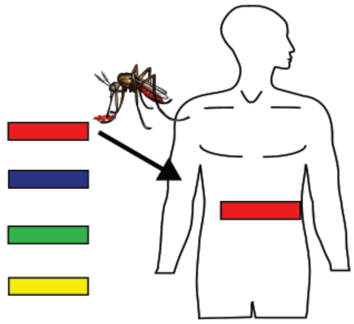
→ impact statistics based on genetic data

Multiplicity of infection - Why is it important?

low transmission



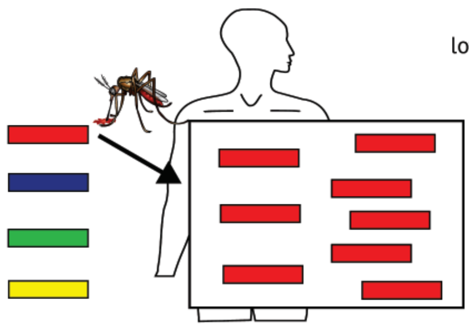
Multiplicity of infection - Why is it important?



low transmission

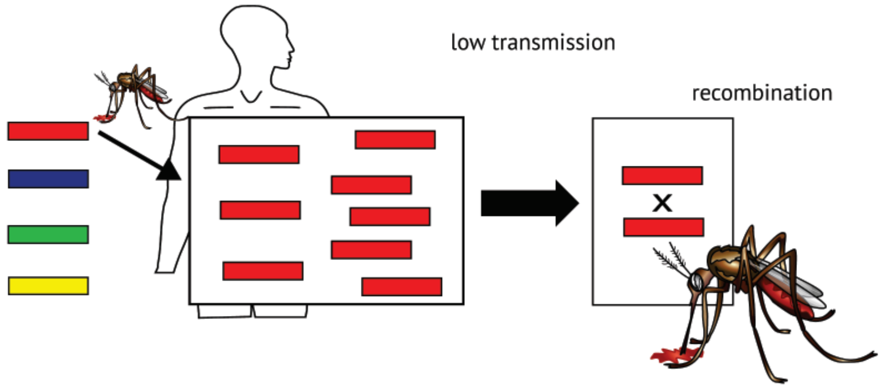
Multiplicity of infection - Why is it important?

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES



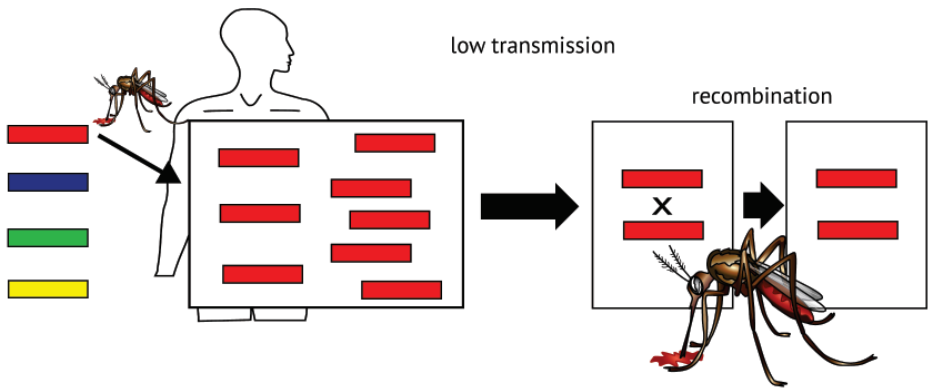
low transmission

Multiplicity of infection - Why is it important?

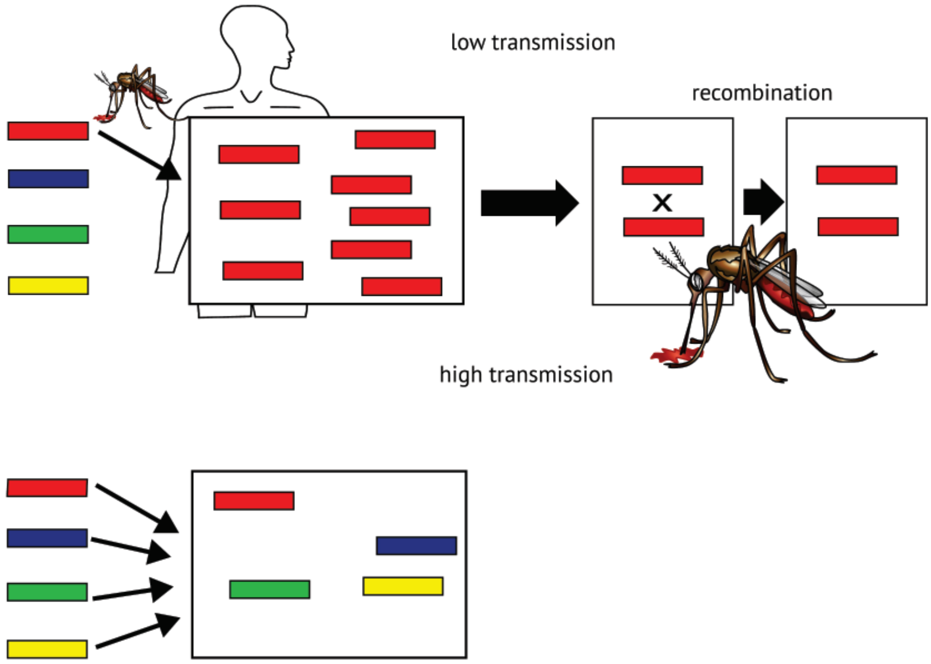


Multiplicity of infection - Why is it important?

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

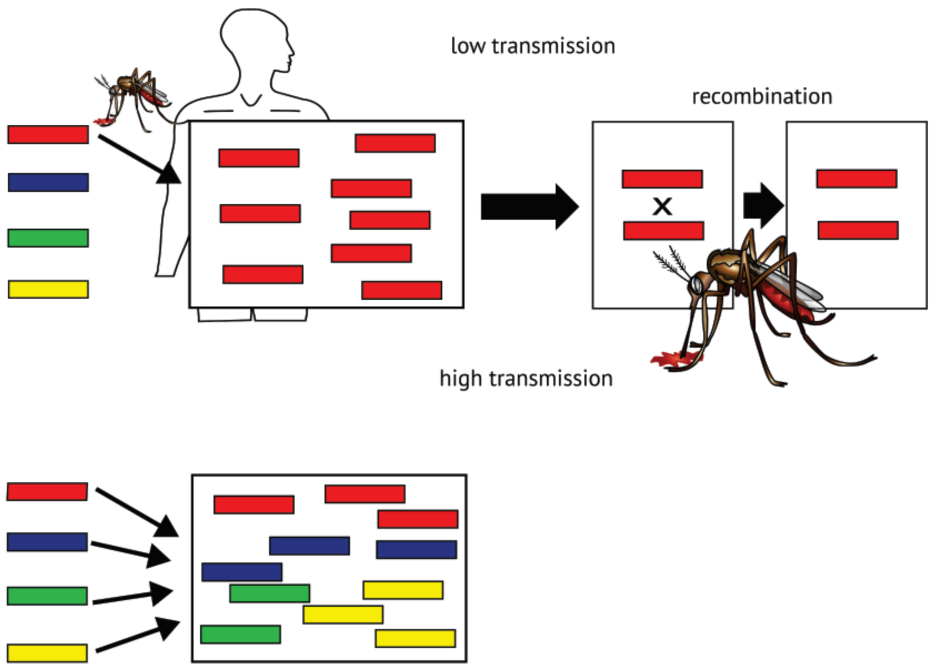


Multiplicity of infection - Why is it important?

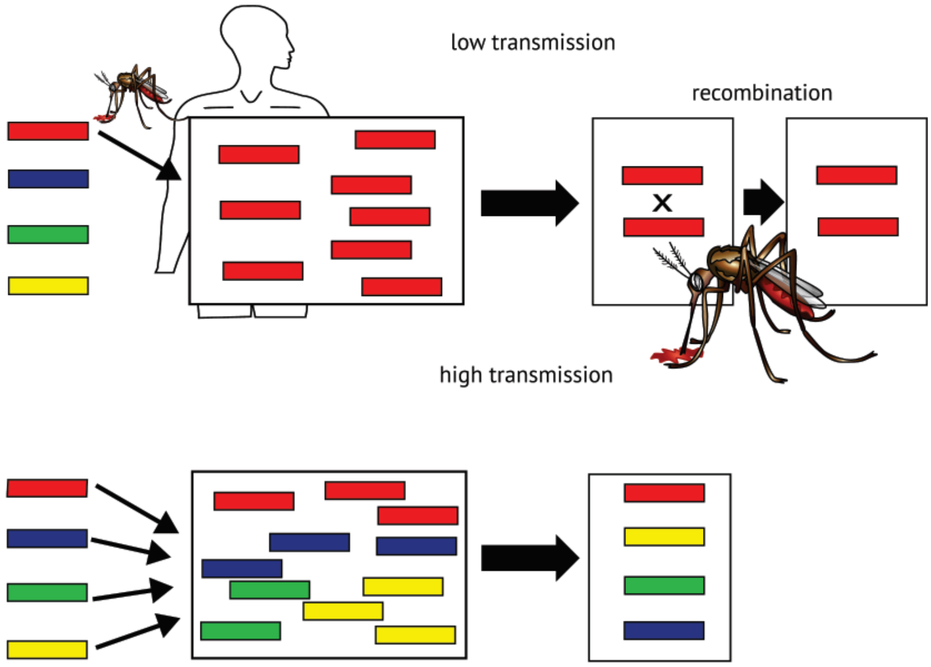


Multiplicity of infection - Why is it important?

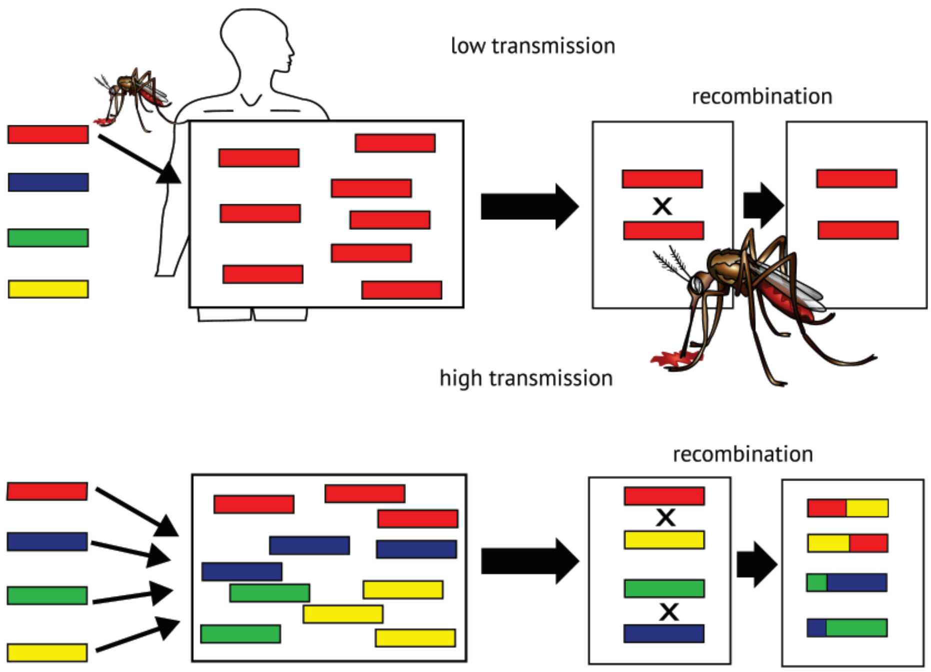
HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

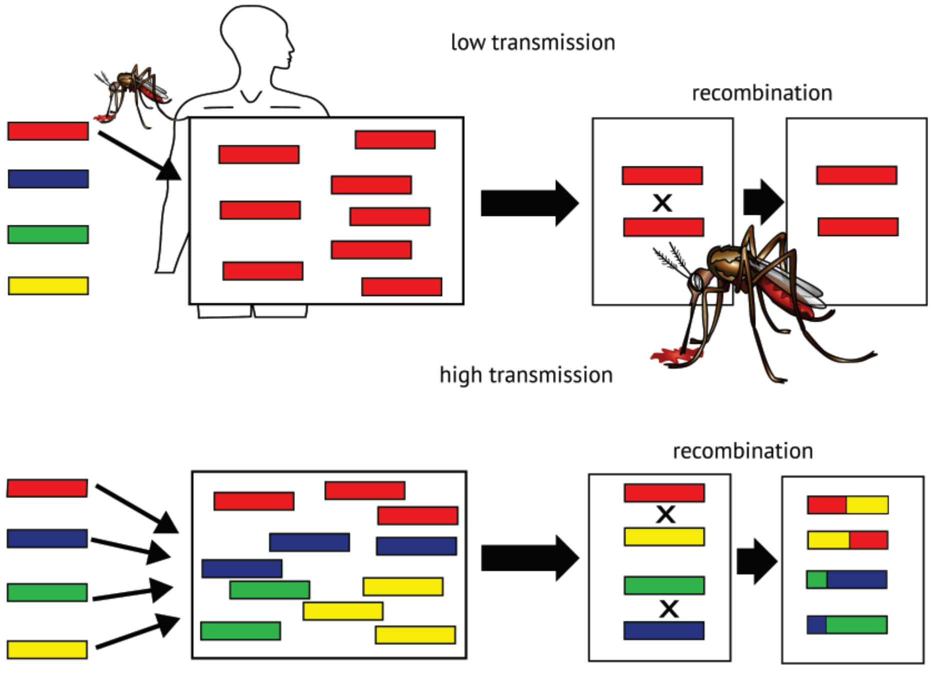


Multiplicity of infection - Why is it important?



Multiplicity of infection - Why is it important?

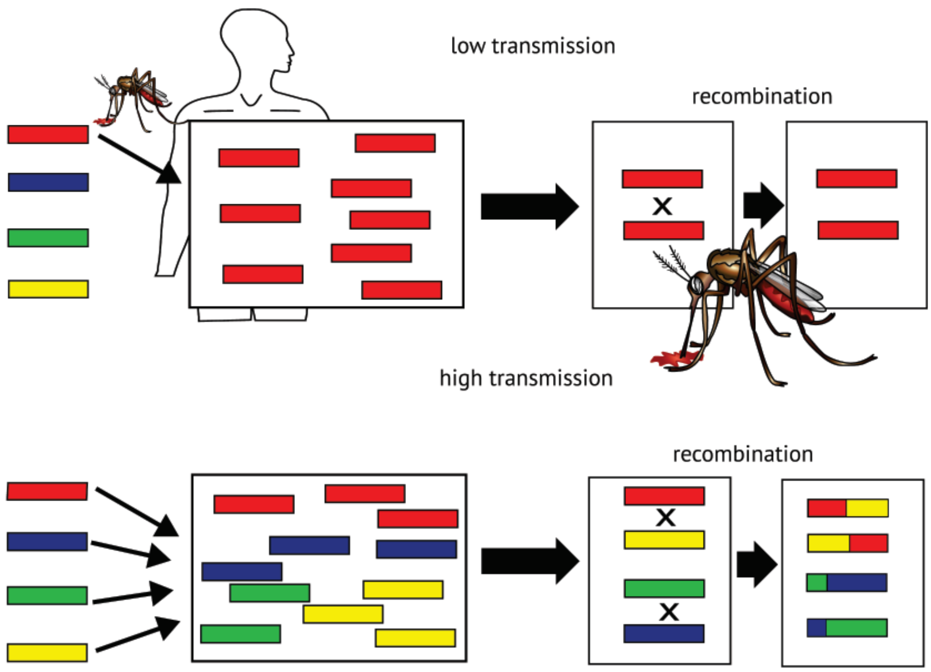




- Large number of co-infections = high transmission

Multiplicity of infection - Why is it important?

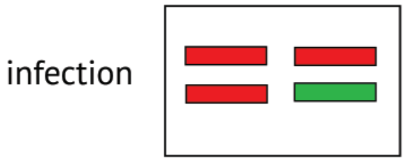
HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES



- Large number of co-infections = high transmission
- High transmission = more genetic variation

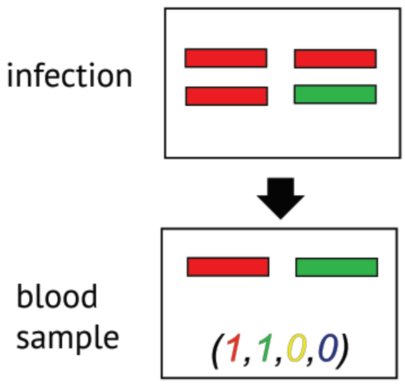
  → Multiplicity of infection = key quantity in genetic studies

Approach

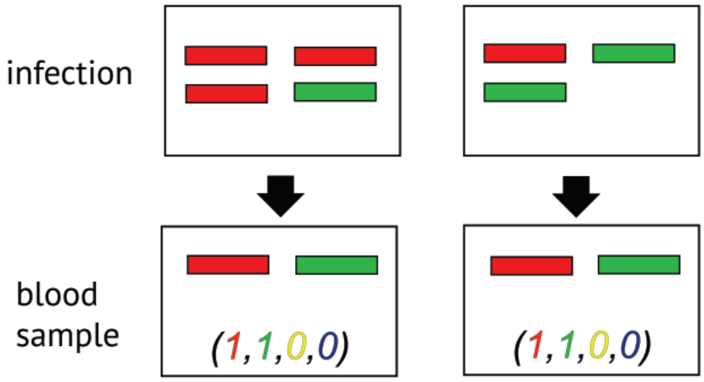


HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

Approach

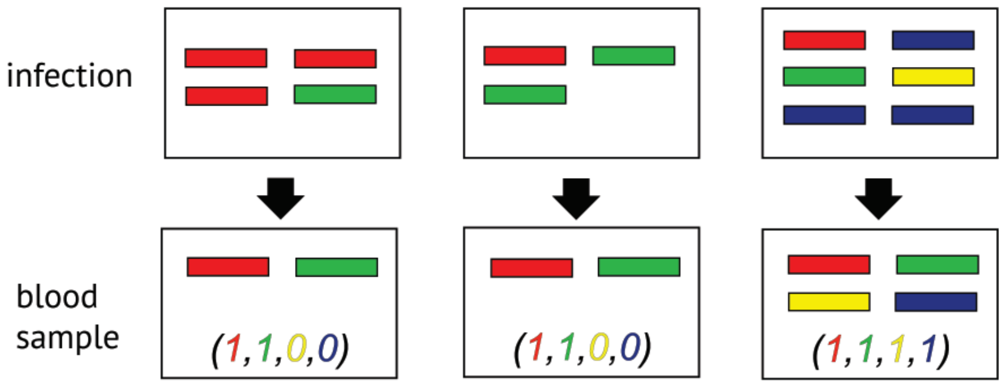


Approach

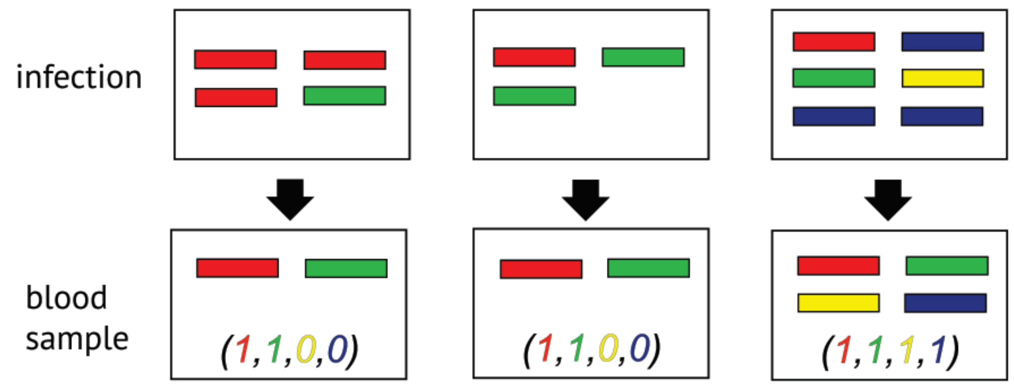


- N blood samples

Approach



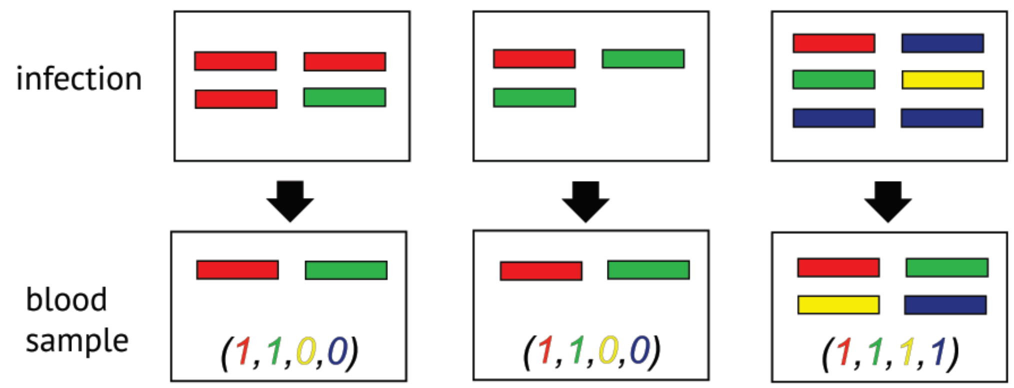
- N blood samples
- n strains (versions of a genetic markers)



- N blood samples
- n strains (versions of a genetic markers)
- n_i observed number samples with configuration $\boldsymbol{i}$ (0-1 vector; length n)

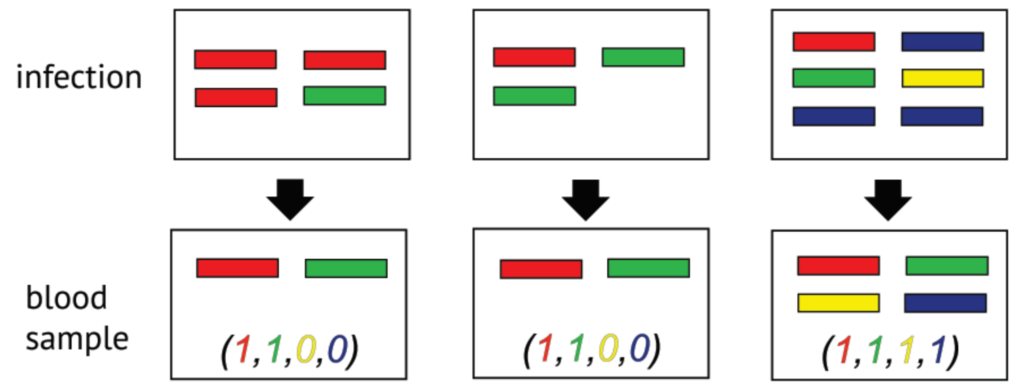
Approach

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES



- N blood samples
- n strains (versions of a genetic markers)
- n_i observed number samples with configuration $\boldsymbol{i}$ (0-1 vector; length n)
- Q_i expected frequency samples with configuration $\boldsymbol{i}$

Approach



- N blood samples
- n strains (versions of a genetic markers)
- n_i observed number samples with configuration $\boldsymbol{i}$ (0-1 vector; length n)
- Q_i expected frequency samples with configuration $\boldsymbol{i}$

Likelihood: $\prod_i Q_i^{n_i}$ Log-likelihood: $L = \sum_i n_i \log Q_i$

How to derive Q_i?

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Assumptions:

How to derive Q_i?

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain

How to derive Q_i?

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events

How to derive Q_i?

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events
- Implications:
  - Number of infecting strains $\sim$ positive Poisson distribution (parameter λ)

How to derive Q_i?

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events
- Implications:
  - Number of infecting strains $\sim$ positive Poisson distribution (parameter λ)

$$P(X = m) = \frac{1}{e^{\lambda}-1} \frac{\lambda^m}{m!} \qquad m > 1$$

How to derive Q_i?

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events
- Implications:
  - Number of infecting strains $\sim$ positive Poisson distribution (parameter λ)

    $P(X = m) = \frac{1}{e^{\lambda}-1} \frac{\lambda^m}{m!} \qquad m > 1$
  - Infection conditioned on $X = m$ strains multinomially distributed $\sim \mathrm{Mult}(m, \boldsymbol{p})$

How to derive Q_i?

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events
- Implications:
  - Number of infecting strains $\sim$ positive Poisson distribution (parameter λ)

    $P(X = m) = \frac{1}{e^{\lambda}-1} \frac{\lambda^m}{m!} \qquad m > 1$
  - Infection conditioned on $X = m$ strains multinomially distributed $\sim \mathrm{Mult}(m, \boldsymbol{p})$
  - **Quantity of interest** $= \frac{\lambda}{1-e^{-\lambda}}$ (mean of positive Poisson distr.)

How to derive Q_i?

- Assumptions:
  - $\boldsymbol{p} = (p_1, \ldots, p_n) \ldots$ frequency vector of strain
  - Infections rare & independent events
- Implications:
  - Number of infecting strains $\sim$ positive Poisson distribution (parameter λ)

    $P(X = m) = \frac{1}{e^\lambda - 1} \frac{\lambda^m}{m!} \qquad m > 1$
  - Infection conditioned on $X = m$ strains multinomially distributed $\sim \text{Mult}(m, \boldsymbol{p})$
  - **Quantity of interest** $= \frac{\lambda}{1 - e^{-\lambda}}$ (mean of positive Poisson distr.)
  - Likelihood: $L = L(\lambda, \boldsymbol{p}) = -N \log(e^\lambda - 1) + \sum_{k=1}^{n} N_k \log(e^{\lambda p_k} - 1)$
  - $N_k \ldots$ number of samples with strain k

Results

- Aims:

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)
  - Existence, uniqueness, numerical effort?

Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)
  - Existence, uniqueness, numerical effort?

- Results:
  - $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$ exists and is unique

Results

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)
  - Existence, uniqueness, numerical effort?
- Results:
  - $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$ exists and is unique
  - $\mathbb{E}_{\hat{\boldsymbol{\theta}}}\, N_k = N_k$ expected = observed number of samples containing strain k

Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)
  - Existence, uniqueness, numerical effort?
- Results:
  - $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$ exists and is unique
  - $\mathbb{E}_{\hat{\boldsymbol{\theta}}}\, N_k = N_k$ expected = observed number of samples containing strain k
  - $\hat{\lambda}$ **found by iterating**

$$\lambda_{t+1} = \lambda_t - \frac{\lambda_t + \sum_{k=1}^{n} \log\left(1 - \frac{N_k}{N}(1 - e^{-\lambda_t})\right)}{1 - \sum_{k=1}^{n} \frac{N_k}{Ne^{\lambda_t} - N_k(e^{-\lambda_t} - 1)}},$$

converges from all starting values $\lambda_0 > \hat{\lambda}$

- Aims:
  - ML estimate of parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$?

    $\rightarrow$ multidimensional Newton method (n independent parameters)
  - Existence, uniqueness, numerical effort?
- Results:
  - $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{p}})$ exists and is unique
  - $\mathbb{E}_{\hat{\boldsymbol{\theta}}} N_k = N_k$ expected = observed number of samples containing strain k
  - $\hat{\lambda}$ **found by iterating**

$$\lambda_{t+1} = \lambda_t - \frac{\lambda_t + \sum_{k=1}^{n} \log\left(1 - \frac{N_k}{N}(1 - e^{-\lambda_t})\right)}{1 - \sum_{k=1}^{n} \frac{N_k}{Ne^{\lambda_t} - N_k(e^{-\lambda_t} - 1)}},$$

converges from all starting values $\lambda_0 > \hat{\lambda}$

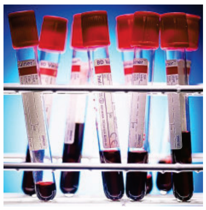
  - $\hat{p}_k = -\frac{1}{\hat{\lambda}} \log\left(1 - \frac{N_k}{N}(1 - e^{-\hat{\lambda}})\right).$

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Cameroon

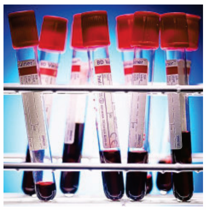
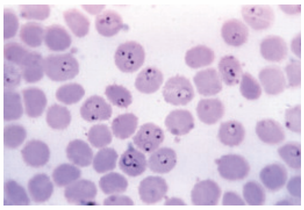
HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Cameroon
- 331 blood sample



Results

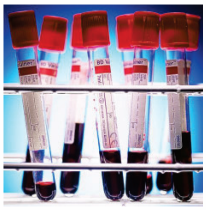
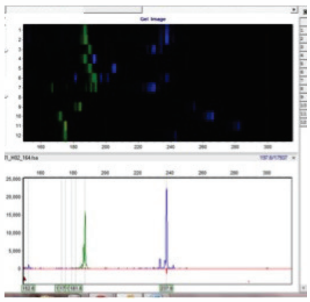
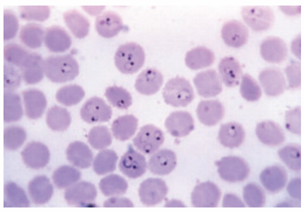
- Cameroon
- 331 blood sample

Results

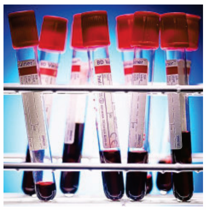
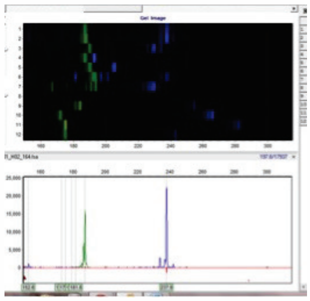
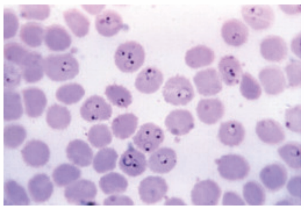
HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Cameroon
- 331 blood sample

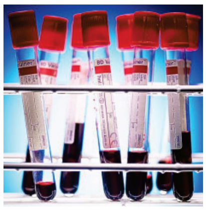
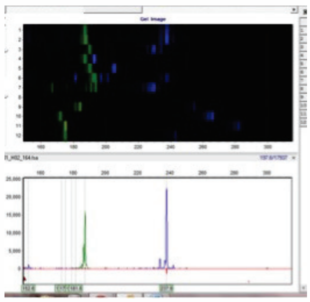
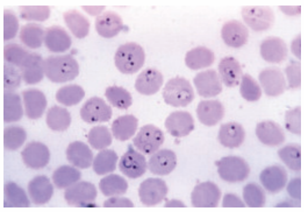
Results

- Cameroon
- 331 blood sample

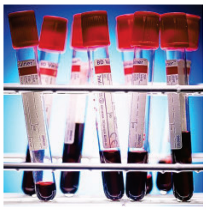





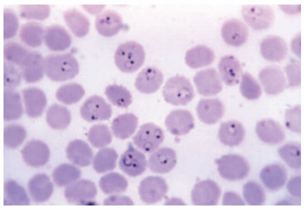

Results

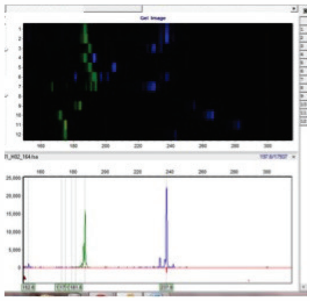
- Cameroon
- 331 blood sample


- Cameroon
- 331 blood sample
- 8 microsatellite markers



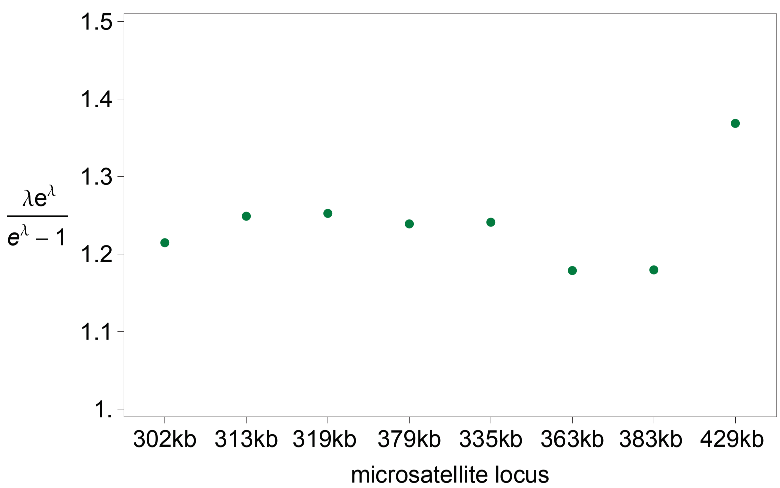




Results

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

- Cameroon
- 331 blood sample
- 8 microsatellite markers



- Aims:
  - Confidence intervalls

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - Confidence intervalls
  - Testing the estimates

- Aims:
  - Confidence intervalls
  - Testing the estimates
- Approach: **Profile Likelihood**
  - λ fixed $\quad L(\lambda, \boldsymbol{p}) \rightarrow \max$?

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - Confidence intervalls
  - Testing the estimates
- Approach: **Profile Likelihood**
  - λ fixed $\quad L(\lambda, \boldsymbol{p}) \rightarrow \max$?
  - $2(\underbrace{\max_{\lambda,\boldsymbol{p}} L(\lambda, \boldsymbol{p})}_{\text{max. LH}} - \underbrace{\max_{\boldsymbol{p}} L(\lambda_0, \boldsymbol{p})}_{\text{profile LH}}) \sim \chi^2_1$

- Aims:
  - Confidence intervalls
  - Testing the estimates
- Approach: **Profile Likelihood**
  - λ fixed $\quad L(\lambda, \boldsymbol{p}) \to \max$?
  - $2(\underbrace{\max_{\lambda,\boldsymbol{p}} L(\lambda, \boldsymbol{p})}_{\text{max. LH}} - \underbrace{\max_{\boldsymbol{p}} L(\lambda_0, \boldsymbol{p})}_{\text{profile LH}}) \sim \chi_1^2$
  - Find $\underline{\lambda}, \overline{\lambda}$ with: $\max_{\boldsymbol{p}} L(\underline{\lambda}, \boldsymbol{p}) = \max_{\boldsymbol{p}} L(\overline{\lambda}, \boldsymbol{p}) = \max_{\lambda,\boldsymbol{p}} L(\lambda, \boldsymbol{p}) - \chi_1^2(1-\alpha)/2 = \ell^*$

Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- Aims:
  - Confidence intervalls
  - Testing the estimates
- Approach: **Profile Likelihood**
  - λ fixed $\quad L(\lambda, \boldsymbol{p}) \to \max$?
  - $2(\underbrace{\max_{\lambda,\boldsymbol{p}} L(\lambda, \boldsymbol{p})}_{\text{max. LH}} - \underbrace{\max_{\boldsymbol{p}} L(\lambda_0, \boldsymbol{p})}_{\text{profile LH}}) \sim \chi_1^2$
  - Find $\underline{\lambda}, \overline{\lambda}$ with: $\max_{\boldsymbol{p}} L(\underline{\lambda}, \boldsymbol{p}) = \max_{\boldsymbol{p}} L(\overline{\lambda}, \boldsymbol{p}) = \max_{\lambda,\boldsymbol{p}} L(\lambda, \boldsymbol{p}) - \chi_1^2(1-\alpha)/2 = \ell^*$
  - Trick: maximize conditioned on $L(\lambda, \boldsymbol{p}) = \ell^*$

    $\to$ Lagrange multiplies & (n+1)-dimensional Newton method

Results:

- $1 - \alpha$ confidence intervals exist & uniquely defined

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

Results:

- $1 - \alpha$ confidence intervals exist & uniquely defined

- Bounds found by iterating 2-dim recursion

Results

Results:

- $1-\alpha$ confidence intervals exist & uniquely defined
- Bounds found by iterating 2-dim recursion
- Approach converges (locally) quadratically

Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

Results:

- $1-\alpha$ confidence intervals exist & uniquely defined
- Bounds found by iterating 2-dim recursion
- Approach converges (locally) quadratically

- Test for $H_0: \lambda = \lambda_0$ vs. $H_A: \lambda \neq \lambda_0$

  reject H_0 if $\lambda_0 \notin [\underline{\lambda}, \overline{\lambda}]$

  p-value:

  $$\chi_1^2\Big(2(\max_{\lambda,\boldsymbol{p}} L(\lambda,\boldsymbol{p}) - \max_{\boldsymbol{p}} L(\lambda_0,\boldsymbol{p}))\Big)$$

Results

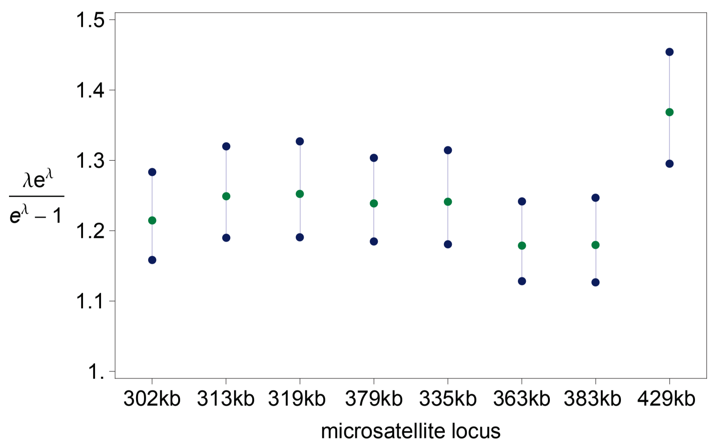
Results:

- 1 − α confidence intervals exist & uniquely defined
- Bounds found by iterating 2-dim recursion
- Approach converges (locally) quadratically
- Test for $H_0: \lambda = \lambda_0$ vs. $H_A: \lambda \neq \lambda_0$

  reject H_0 if $\lambda_0 \notin [\underline{\lambda}, \overline{\lambda}]$

  p-value:

  $$\chi_1^2\Big(2(\max_{\lambda,\boldsymbol{p}} L(\lambda,\boldsymbol{p}) - \max_{\boldsymbol{p}} L(\lambda_0,\boldsymbol{p}))\Big)$$



Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

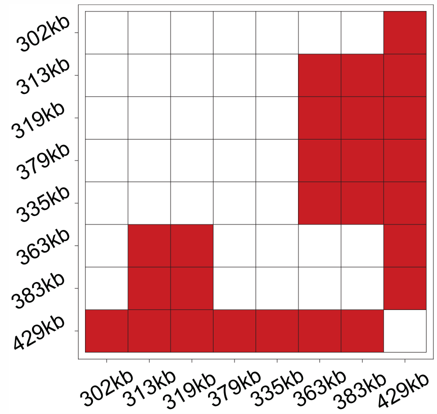
Results:

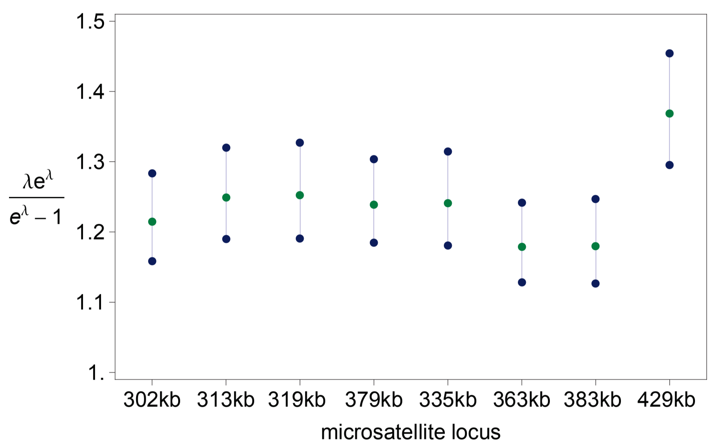
- 1 − α confidence intervals exist & uniquely defined
- Bounds found by iterating 2-dim recursion
- Approach converges (locally) quadratically
- Test for $H_0: \lambda = \lambda_0$ vs. $H_A: \lambda \neq \lambda_0$

  reject H_0 if $\lambda_0 \notin [\underline{\lambda}, \overline{\lambda}]$

  p-value:

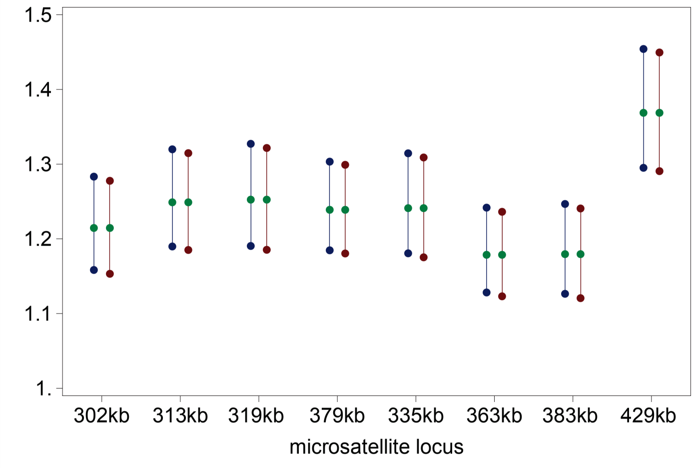
  $$\chi_1^2\Big(2(\max_{\lambda,\boldsymbol{p}} L(\lambda,\boldsymbol{p}) - \max_{\boldsymbol{p}} L(\lambda_0,\boldsymbol{p}))\Big)$$

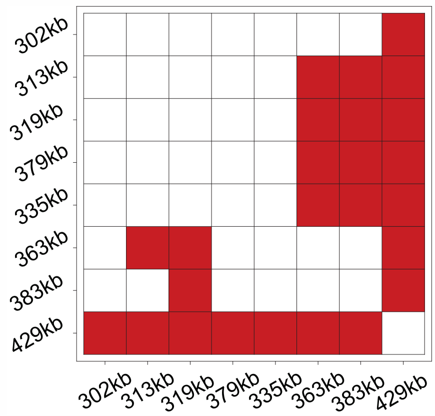


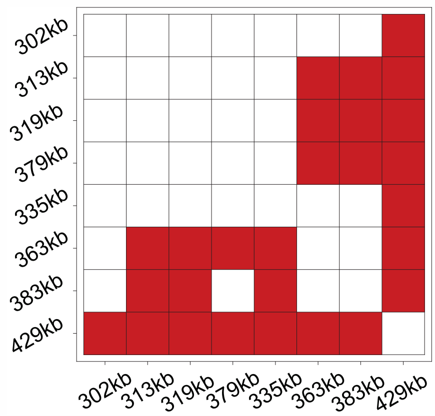


Results

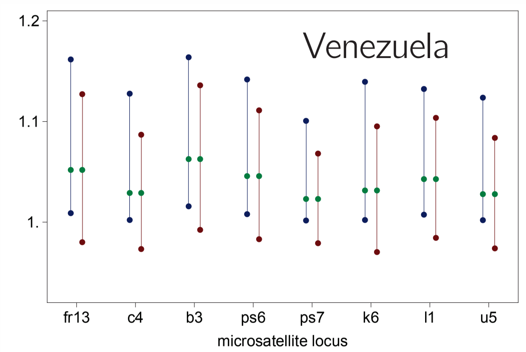
$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, I_N^{-1}(\boldsymbol{\theta}_0))$$

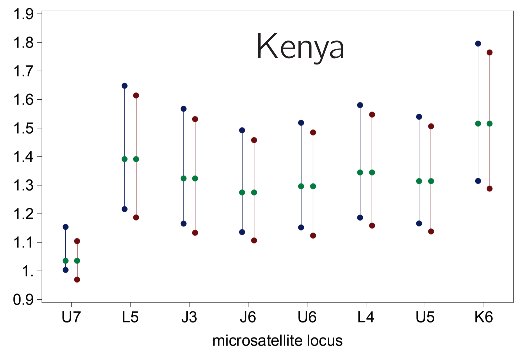






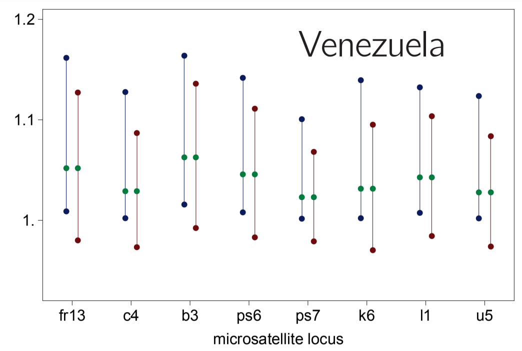
Results

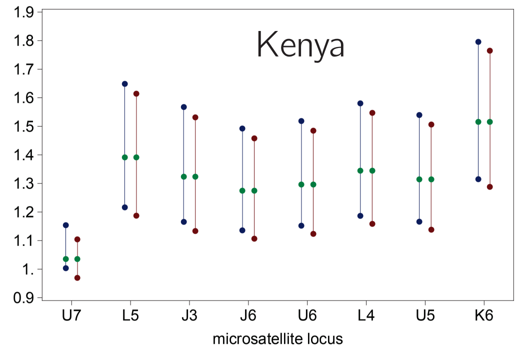


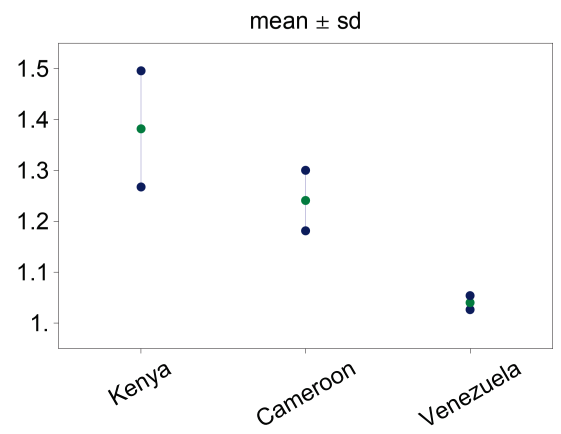


Results

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES







Conclusions

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- ML estimate for number of co-infections

  (key parameter in malaria)

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- ML estimate for number of co-infections

  (key parameter in malaria)
- ML-approach works well

Conclusions

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- ML estimate for number of co-infections

  (key parameter in malaria)
- ML-approach works well
- Robustness study $\rightarrow$ future work

- ML estimate for number of co-infections

  (key parameter in malaria)
- ML-approach works well
- Robustness study → future work
- Including several markers at the same time

Conclusions

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- ML estimate for number of co-infections

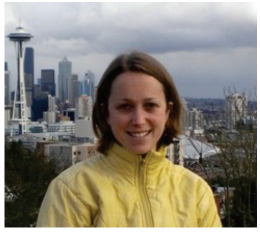
  (key parameter in malaria)
- ML-approach works well
- Robustness study → future work
- Including several markers at the same time

McCollum et al, 2012
Malaria J
Schneider & Kim, 2010,
Theo. Pop. Biol.

HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

- ML estimate for number of co-infections

  (key parameter in malaria)
- ML-approach works well
- Robustness study → future work
- Including several markers at the same time



McCollum et al, 2012
Malaria J
Schneider & Kim, 2010,
Theo. Pop. Biol.