

# Příklady vyhodnocení evolučních algoritmů

---

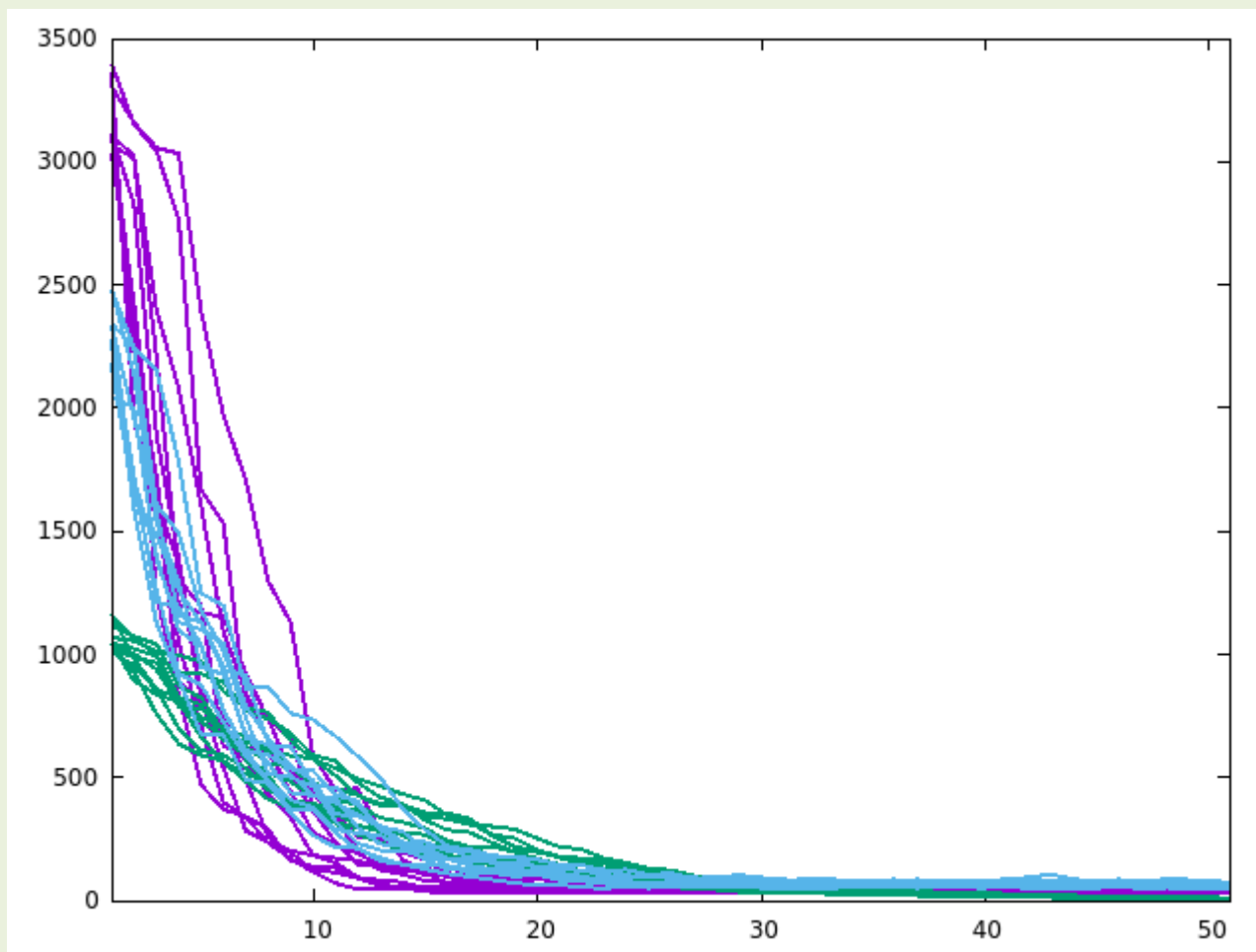
*Martin Hyrš*

*Fakulta informačních technologií*

*Vysoké učení technické v Brně*

*ihyrs@fit.vutbr.cz*

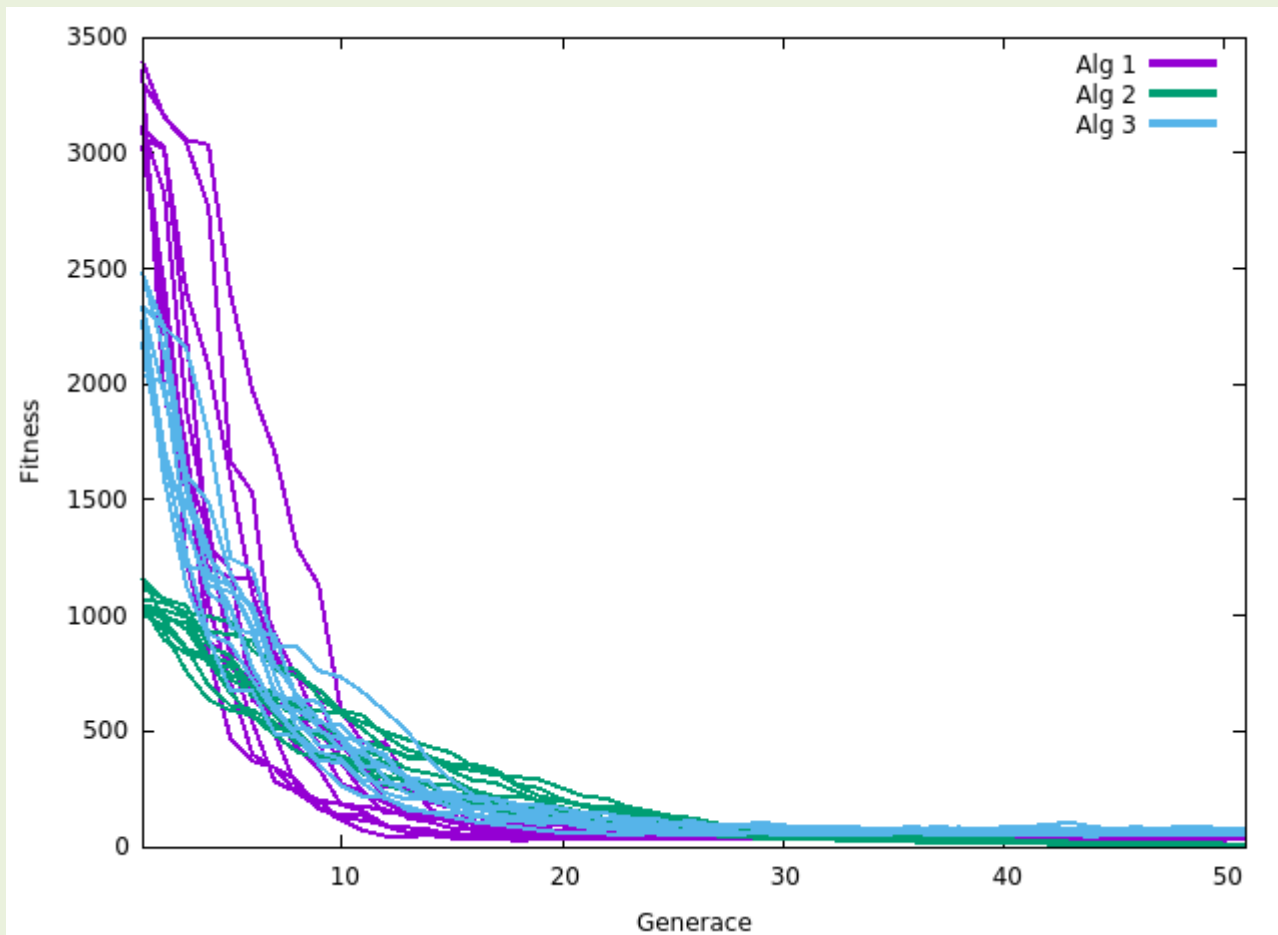
# Opakované běhy algoritmu



☑ Opakované běhy EA jsou při vyhodnocování nutnost, jeden běh nevypovídá o ničem.

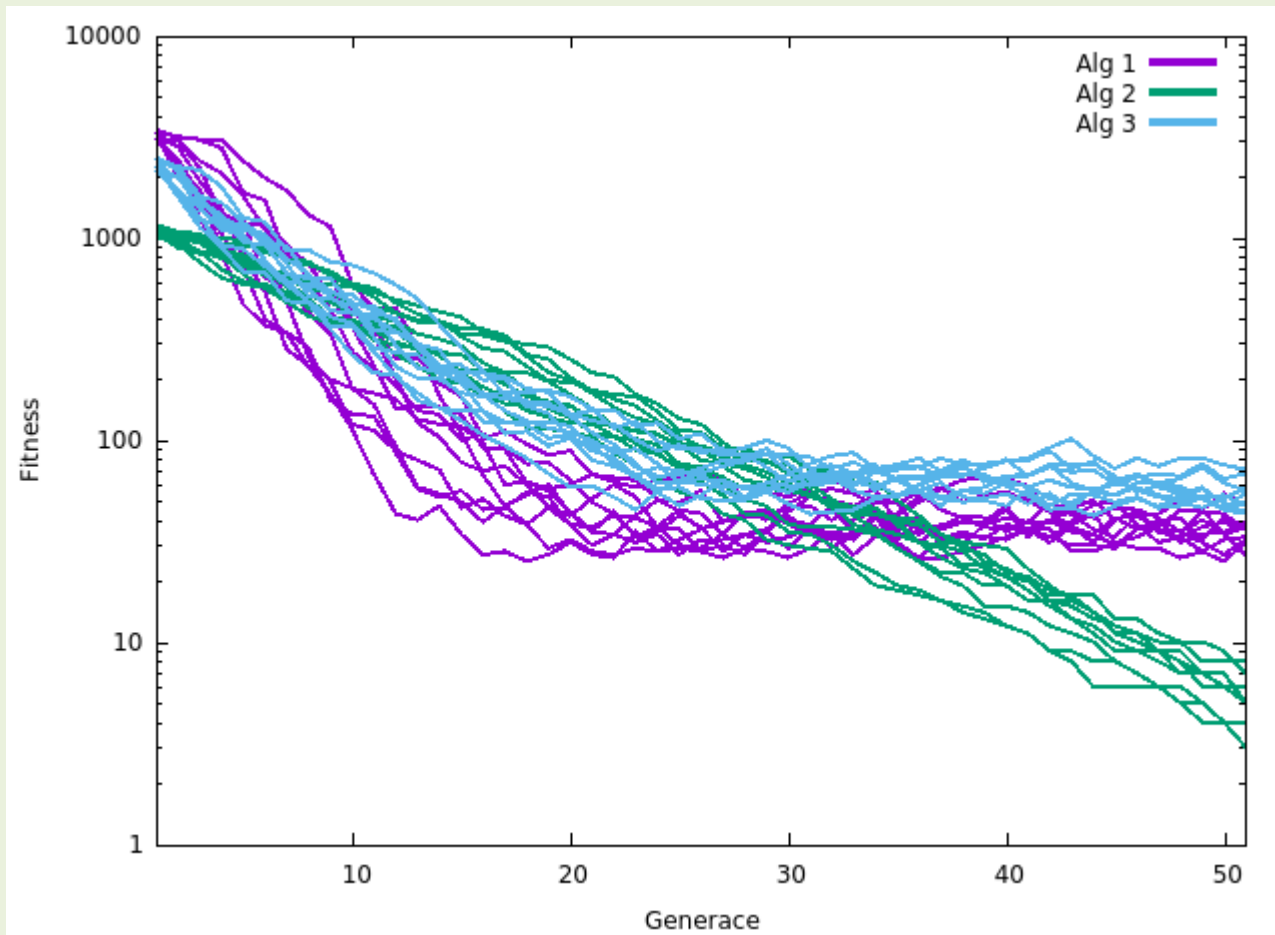
☒ Co je na grafu?

# Popsat graf



☒ U pozdějších generací nerozeznáme podrobnosti.  
(Je samozřejmé, že na začátku jsou všechny algoritmy špatné. Ale který je nejlepší ve 30. generaci?)

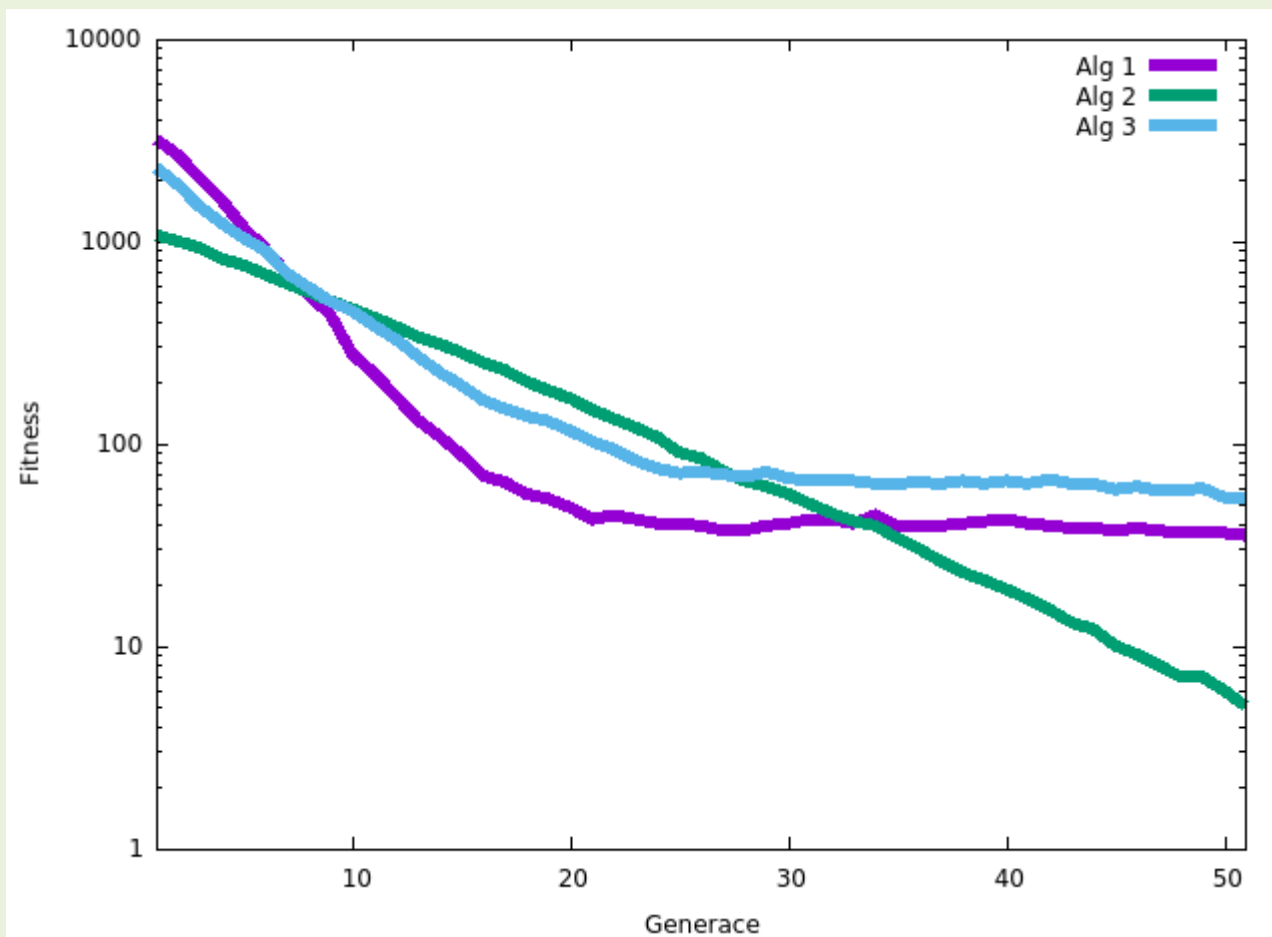
# Nebude logaritmické měřítko přehlednější?



☑ Možná bude, možná nebude; to závisí na datech.

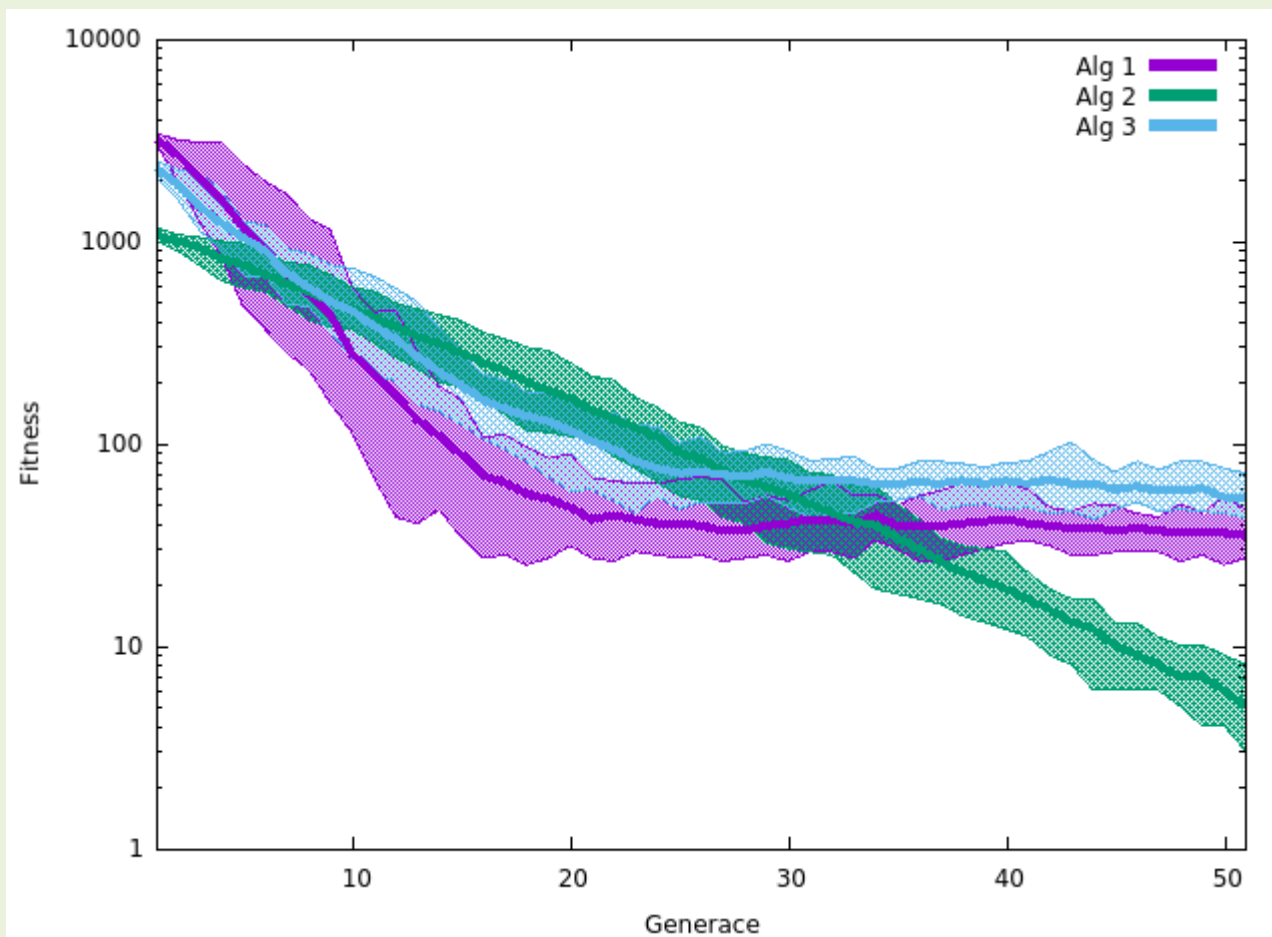
Ale stále tam je moc čar.

# Jen průměr



✗ Málo informací, nevíme, jak se liší různé běhy.

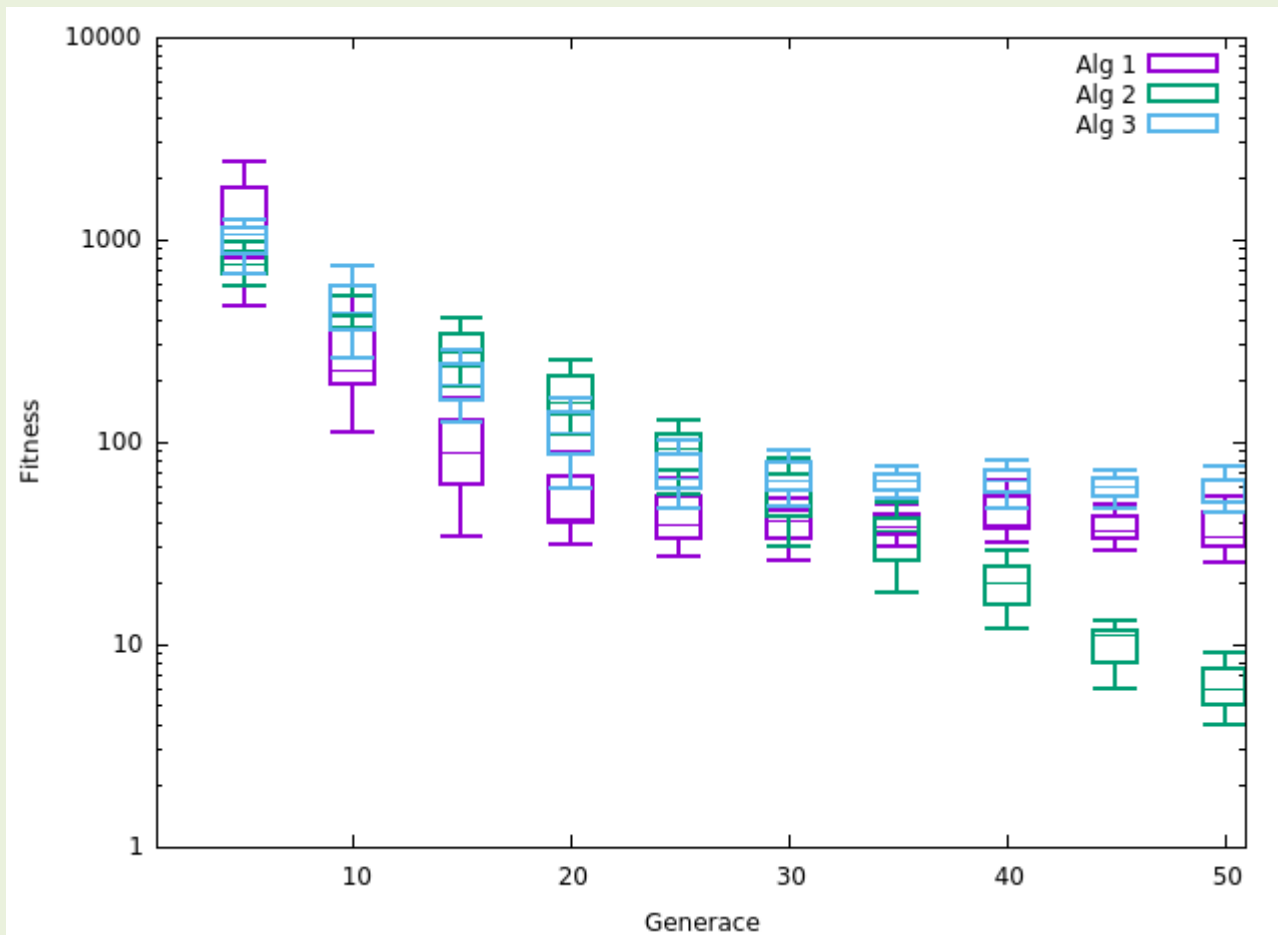
# Vyplněné křivky



Vyplnit plochu mezi min a max, zvýraznit průměr.

☒ Použitelné provedení.

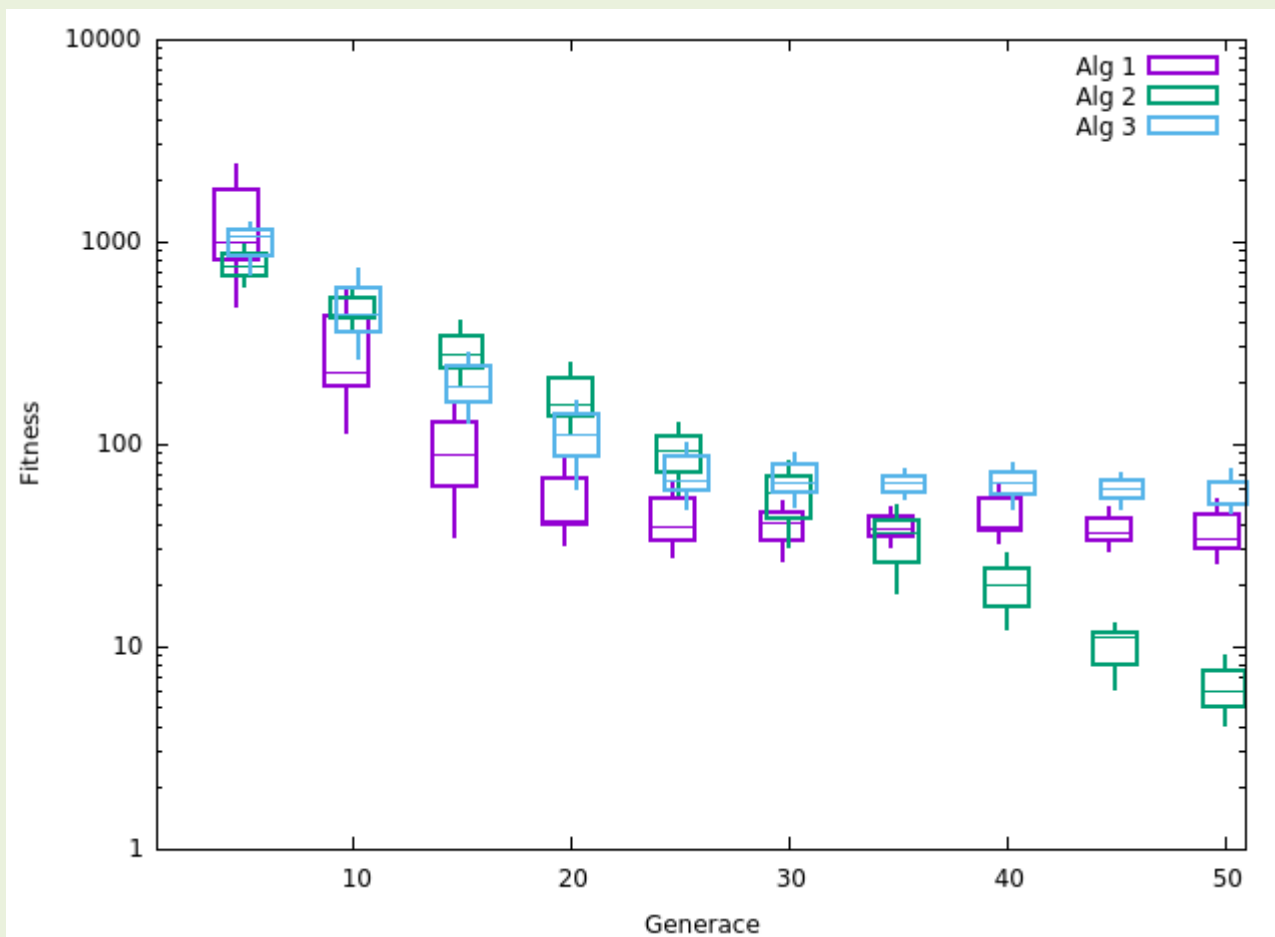
# Boxplot



min — **Q1 MED Q3** — max

Moderní, dostatek informací. Při větším počtu nepřehledné.

# Zjednodušení boxplotů

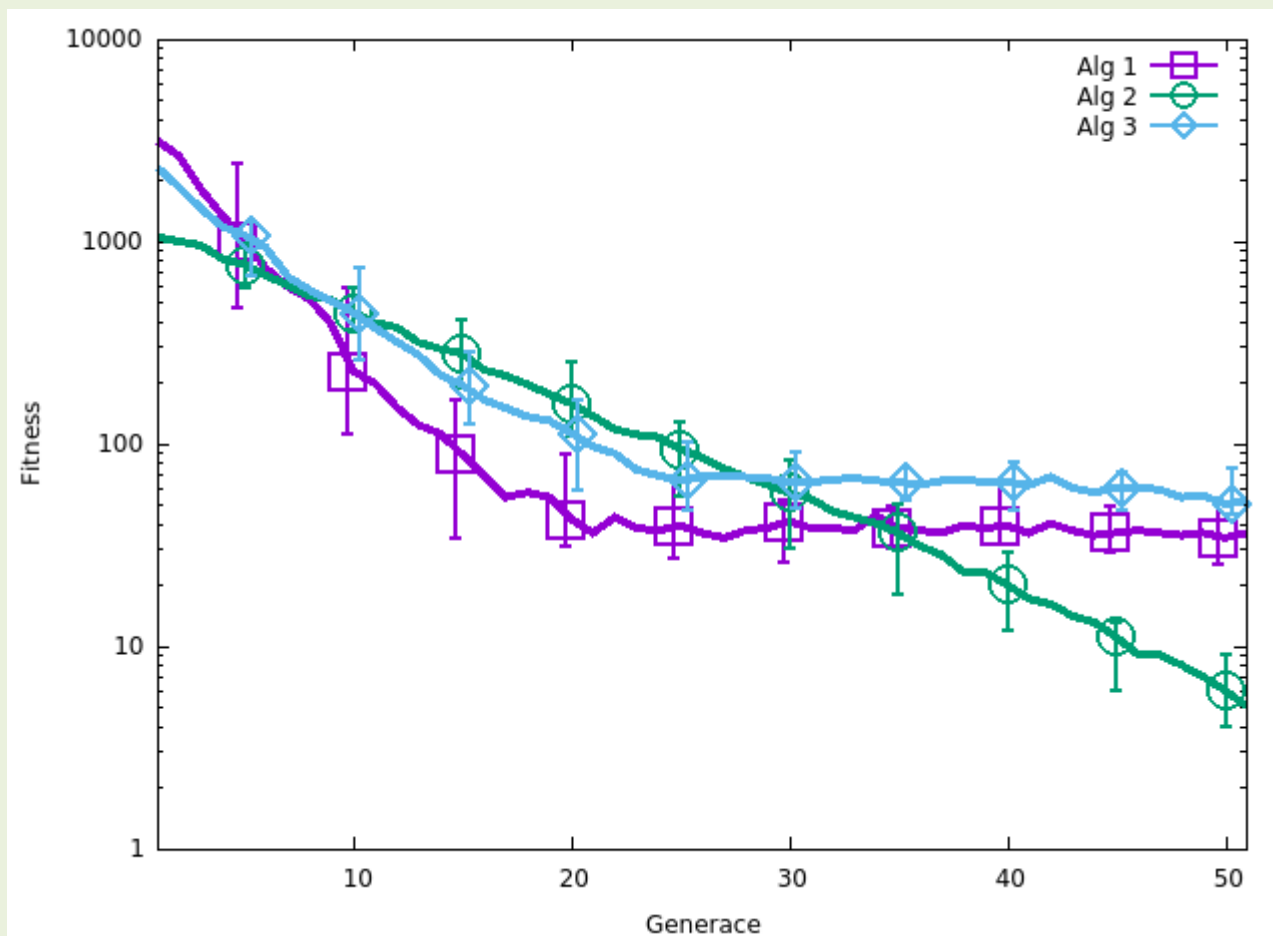


Mírný vodorovný posun, chybí koncové příčky.

☑ Možná přehlednější?

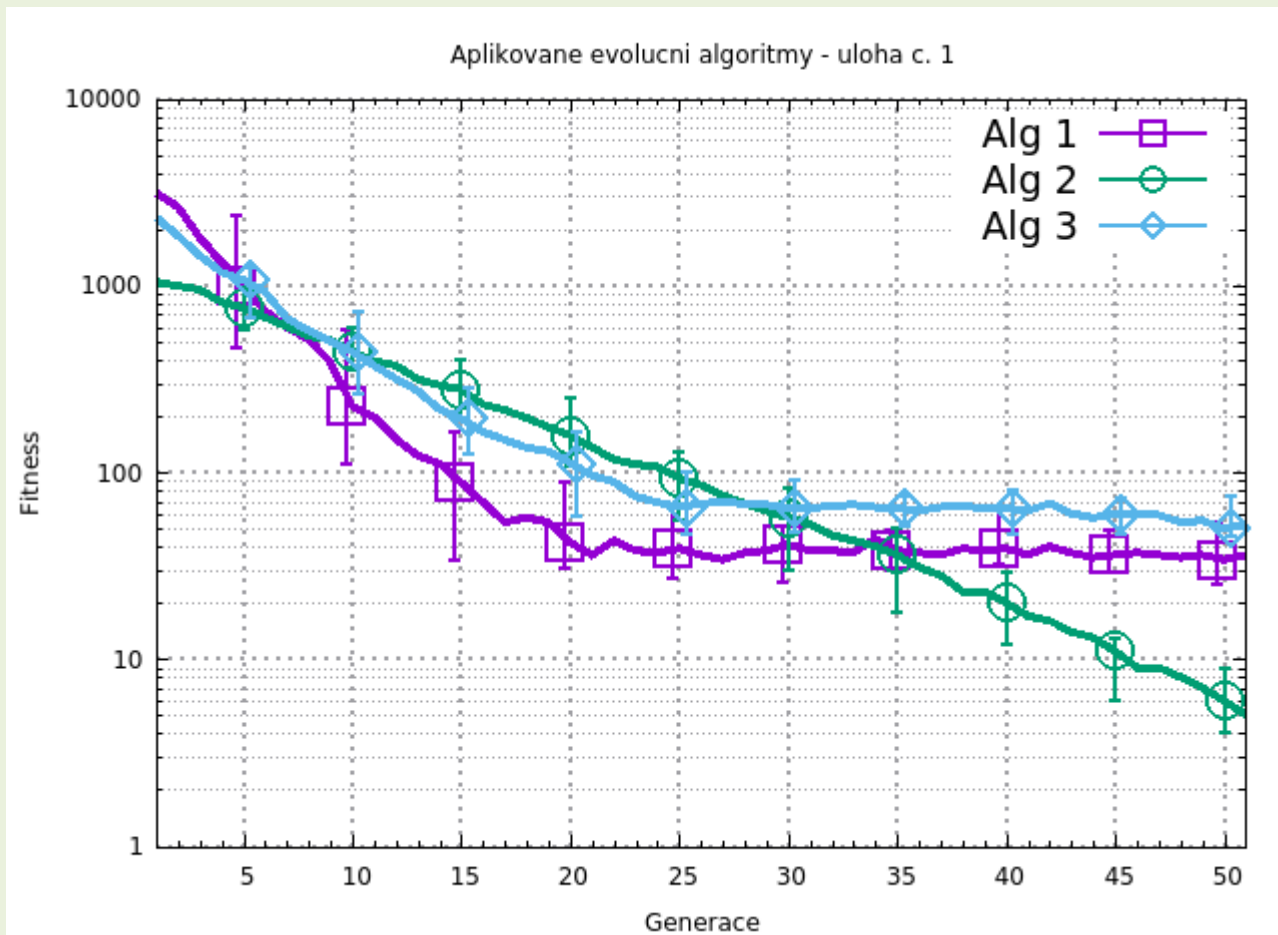


# Čára s body – medián, vousy – min/max



Jednotlivé algoritmy můžeme rozlišit barvou čáry a/nebo typem bodu. Při vhodném provedení lze zobrazit i mnoho algoritmů naráz.

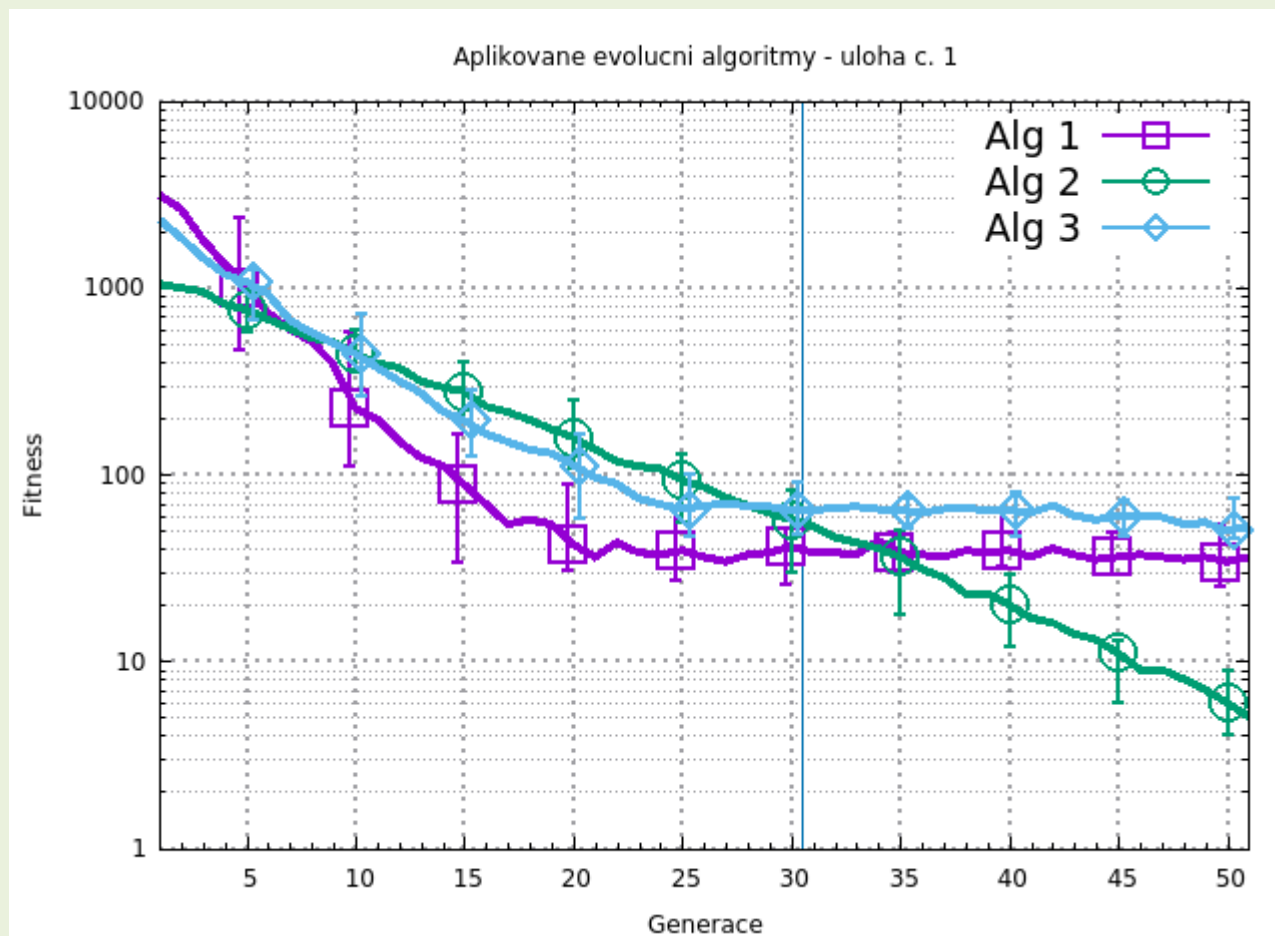
# Další vylepšení



☑ popisky, mřížka, značky na osách...

Je to už dokonalé?

# Co vlastně měříme ???



- Který algoritmus je nejlepší po 30 generacích?
- Stačí dosažení hodnoty menší než 100?

# Možné metriky (např.:)

Po stanoveném počtu vyhodnocení fitness (spravedlivější než počet generací, když jsou různě velké populace):

- úspěšnost (success rate) – při kolika % běhů bylo dosaženo stanoveného prahu.
- fitness – je-li menší než práh, považujeme za 0. Statistiky o dosažené fitness.

TABULKA 1: Fitness hodnota nejlepšího jedince po uplynutí 30 generací, statistika z 10 nezávislých běhů na zkušební úloze č. 1.

| Alg.  | min | med | max | avg  | std  |
|-------|-----|-----|-----|------|------|
| Alg 1 | 29  | 38  | 58  | 42,7 | 9,5  |
| Alg 2 | 29  | 54  | 72  | 50,5 | 12,4 |
| Alg 3 | 43  | 65  | 83  | 65,6 | 10,3 |

# Testování hypotéz

---

Chceme porovnat, zda dvě náhodné veličiny mají stejnou střední hodnotu.

$H_0$  – ano, jsou stejné, pozorovaný rozdíl je způsoben pouze náhodou.

$H_1$  – mezi zkoumanými veličinami je významný rozdíl.

**Chyba I. nebo II. druhu:** zamítnutí platného nebo přijetí neplatného.

**Hladina významnosti  $\alpha$**  = pravděpodobnost chyby I. druhu. Obvykle se volí  $\alpha = 5 \%$ .

# Dvouvýběrový t-test

---

Testuje, zda dva nezávislé náhodné výběry mají stejné střední hodnoty. (Za předpokladu, že oba výběry mají stejný počet vzorků. Jinak se musí nejprve rozhodnout F-testem, zda mají stejné rozptyly, a dále zvolit příslušnou variantu t-testu.)

Z průměrů a rozptylů obou výběrů a počtu vzorků vypočítáme testovací kritérium  $t$ :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{N}}}$$

Vypočtenou statistiku  $t$  porovnáme s tabulkovou kritickou hodnotou  $t_{1-\alpha/2}(2N - 2)$ .

Pokud je  $t > t_{1-\alpha/2}$ , je rozdíl mezi výběry statisticky významný na zvolené hladině významnosti  $\alpha$ .

# Statistická významnost rozdílu

Po 30. generaci:

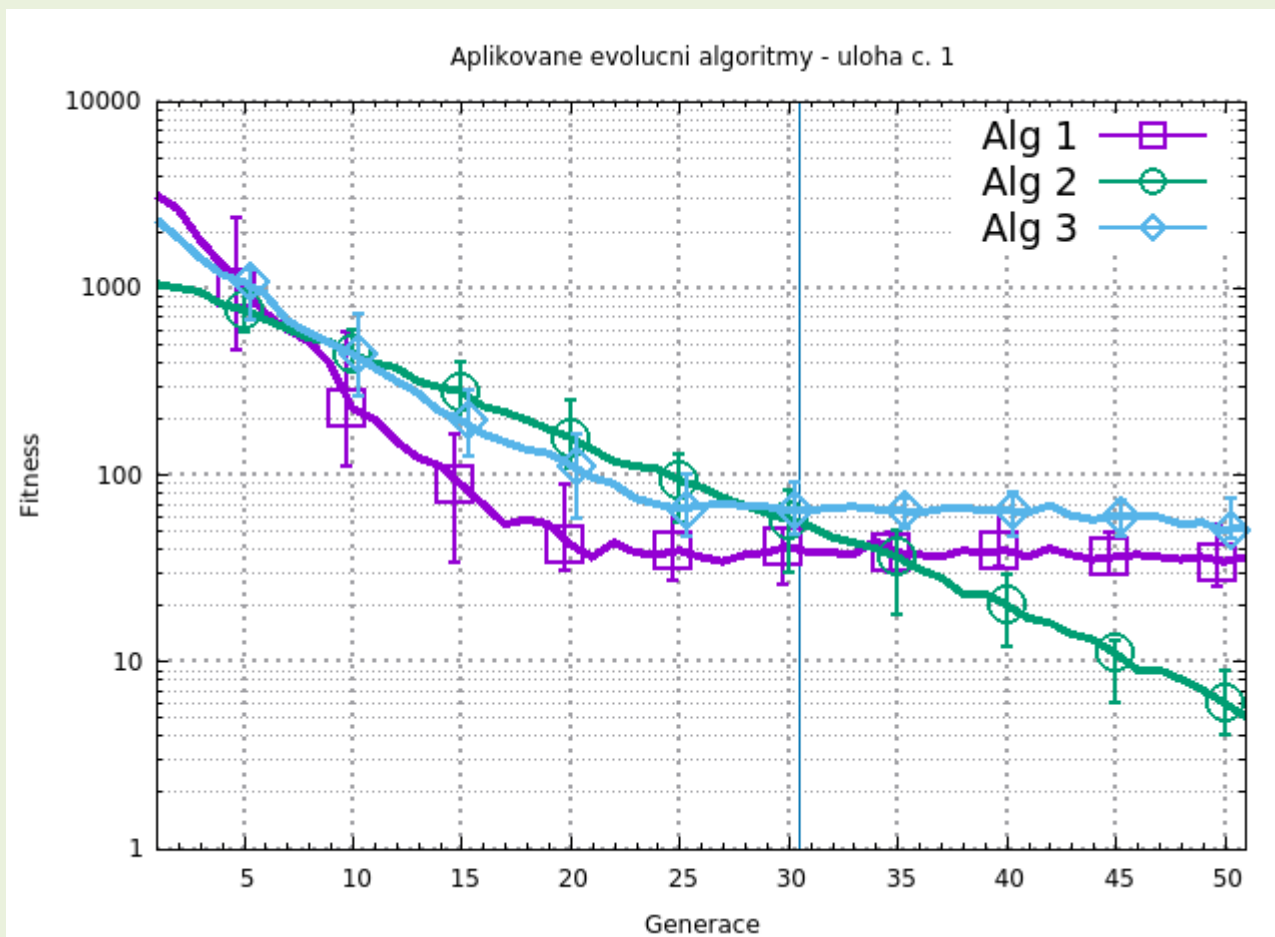
$$\alpha = 0,05 \quad t_{0,975}(18) = 2,101$$

| Alg 1 |     | Alg 2 |      | $t = \frac{ 42,7 - 50,5 }{\sqrt{\frac{9,5^2 + 12,4^2}{10}}} = 1,585$ |
|-------|-----|-------|------|--|
| avg   | std | avg   | std  |  |
| 42,7  | 9,5 | 50,5  | 12,4 |  |

| Alg 1 |     | Alg 3 |      | $t = \frac{ 42,7 - 65,6 }{\sqrt{\frac{9,5^2 + 10,3^2}{10}}} = 5,180$ |
|-------|-----|-------|------|--|
| avg   | std | avg   | std  |  |
| 42,7  | 9,5 | 65,6  | 10,3 |  |

| Alg 2 |      | Alg 3 |      | $t = \frac{ 50,5 - 65,5 }{\sqrt{\frac{12,4^2 + 10,3^2}{10}}} = 2,973$ |
|-------|------|-------|------|---|
| avg   | avg  | avg   | std  |   |
| 50,5  | 12,4 | 65,6  | 10,3 |   |

# Výsledky testu (generace 30)



⇒ Alg. 1 je významně lepší než Alg. 3

⇒ Alg. 2 je významně lepší než Alg. 3

⇒ není významný rozdíl mezi Alg. 1 a Alg. 2



# PS:

---

```
set logscale y
set xlabel "Generace"
set ylabel "Fitness"
set xrange [1:51]

plot for [i=2:11] 'c.txt' u i w lines lc 1 lw 2 t '', \
for [i=2:11] 'b.txt' u i w lines lc 2 lw 2 t '', \
for [i=2:11] 'a.txt' u i w lines lc 3 lw 2 t '', \
1/0 t 'Alg 1' lc 1 lw 4, \
1/0 t 'Alg 2' lc 2 lw 4, \
1/0 t 'Alg 3' lc 3 lw 4

plot for [i=15:15] 'c.txt' u i w lines lc 1 lw 6 t '', \
for [i=15:15] 'b.txt' u i w lines lc 2 lw 6 t '', \
for [i=15:15] 'a.txt' u i w lines lc 3 lw 6 t '', \
1/0 t 'Alg 1' lc 1 lw 6, \
1/0 t 'Alg 2' lc 2 lw 6, \
1/0 t 'Alg 3' lc 3 lw 6

plot 'c.txt' u 0:12:13 w filledcu lc 1 t '' fs pattern 6, \
'b.txt' u 0:12:13 w filledcu lc 2 t '' fs pattern 2, \
'a.txt' u 0:12:13 w filledcu lc 3 t '' fs pattern 7, \
for [i=12:13] 'c.txt' u i w lines lc 1 lw 1 t '', \
for [i=12:13] 'b.txt' u i w lines lc 2 lw 1 t '', \
for [i=12:13] 'a.txt' u i w lines lc 3 lw 1 t '', \
for [i=15:15] 'c.txt' u i w lines lc 1 lw 4 t '', \
for [i=15:15] 'b.txt' u i w lines lc 2 lw 4 t '', \
for [i=15:15] 'a.txt' u i w lines lc 3 lw 4 t '', \
1/0 t 'Alg 1' lc 1 lw 6, \
1/0 t 'Alg 2' lc 2 lw 6, \
1/0 t 'Alg 3' lc 3 lw 6

set boxwidth 2

plot 'c.txt' every 5 u ($1+1):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 1 whiskerbars, \ # Kvartily nejsou skutečné, musí se stanovit z primárních výsledků experimentů.
'' every 5 u ($1+1):14:14:14:14 with candlesticks t '' lw 1 lc 1, \
'b.txt' every 5 u ($1+1):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 2 whiskerbars, \ # Toto je jen příklad pro názornost a proto je kvůli zjednodušení přípravy
'' every 5 u ($1+1):14:14:14:14 with candlesticks t '' lw 1 lc 2, \
'a.txt' every 5 u ($1+1):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 3 whiskerbars, \ # ořízený.
'' every 5 u ($1+1):14:14:14:14 with candlesticks t '' lw 1 lc 3, \
1/0 t 'Alg 1' lw 2 lc 1 w boxes, \
1/0 t 'Alg 2' lw 2 lc 2 w boxes, \
1/0 t 'Alg 3' lw 2 lc 3 w boxes

plot 'c.txt' every 5 u ($1+0.7):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 1, \
'' every 5 u ($1+0.7):14:14:14:14 with candlesticks t '' lw 1 lc 1, \
'b.txt' every 5 u ($1+1):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 2, \
'' every 5 u ($1+1):14:14:14:14 with candlesticks t '' lw 1 lc 2, \
'a.txt' every 5 u ($1+1.3):($15+$12)/2:12:13:($15+$13)/2 with candlesticks t '' lw 2 lc 3, \
'' every 5 u ($1+1.3):14:14:14:14 with candlesticks t '' lw 1 lc 3, \
1/0 t 'Alg 1' lw 2 lc 1 w boxes, \
1/0 t 'Alg 2' lw 2 lc 2 w boxes, \
1/0 t 'Alg 3' lw 2 lc 3 w boxes

plot 'c.txt' u ($1+1):14 w lines lc 1 lw 4 t '', \
'' every 5 u ($1+0.7):14:12:13 w yerrorbars lc 1 lw 2 pt 4 ps 3 t '', \
'b.txt' u ($1+1):14 w lines lc 2 lw 4 t '', \
'' every 5 u ($1+1.0):14:12:13 w yerrorbars lc 2 lw 2 pt 6 ps 3 t '', \
'a.txt' u ($1+1):14 w lines lc 3 lw 4 t '', \
'' every 5 u ($1+1.3):14:12:13 w yerrorbars lc 3 lw 2 pt 12 ps 3 t '', \
1/0 w linespoint lw 2 lc 1 pt 4 ps 2 t 'Alg 1', \
1/0 w linespoint lw 2 lc 2 pt 6 ps 2 t 'Alg 2', \
1/0 w linespoint lw 2 lc 3 pt 12 ps 2 t 'Alg 3'

set grid xtics ytics mytics lw 2, lw 1
set xtics 5
set mxtics 5
set key font ",14"
set title "Aplikovane evolucionni algoritmy - uloha c. 1"
replot

set arrow from 30.5,1 to 30.5,10000 nohead lc 6
replot
```