Delineating fine scale population genetic structure in the fire salamander (*Salamandra salamandra*)

Sermier Roberto

Master thesis 2007

Director: Luca Fumagalli

Supervisors: Glenn Yannick & Helfer Véronique

Laboratoire de biologie de la conservation (LBC)

UNIL | Université de Lausanne
Département d'écologie et évolution

laboratoire de biologie de la conservation
laboratory for conservation biology

Abstract:

Amphibians are currently facing major threat at a global scale. One of the major causes of this decline is anthropogenic driven habitat loss and fragmentation. These factors increase the risk of population extinction. Dispersal plays an important role in this regard as it is linked to patches recolonization rate but also a diminution of inbreeding. As connectivity varies a lot between amphibian species studying connectivity independently for each species appear necessary. In this study, genetic analyses were used to: first investigate the connectivity of *Salamandra salamandra* in canton of Luzern and second try to define a proper conservation unit for this specie. This species is listed by the IUCN as a least concern species. However reduce in its sightings have been observed in Switzerland. Genetic analysis revealed moderate overall isolation (FST = 0.099) between the sampling units. This value ranges with isolation values of similar widespread and unthreatened species. Bayesian analysis revealed most of the sampling units (14 out of 19) could be grouped in two coherent clusters. Hierarchical analysis revealed that most of the structure is sustained by the cluster (F_{ST} = 0.093) but that there still was moderate significant structure between the sampling units within the clusters (F_{ST} = 0.045). Considering these results we propose to adopt the cluster level as the main conservation unit.

Introduction:

Amphibians are currently facing major threat at a global scale, even more than birds or mammals (Drost and Fellers 1996; Sarkar 1996). In Europe, and globally, one of the

major causes of this decline is anthropogenic driven habitat loss and fragmentation (Stuart, Chanson et al. 2004; Funk, Greene et al. 2005). Habitat loss and fragmentation reduce the size of and isolate populations, increasing the risks due to demographic, stochastic and genetic events. These two factors tend to drive populations toward extinction (Gilpin and Soulé 1986). High fragmentation may also result in an increased inbreeding within populations which may lead to inbreeding depression. Inbreeding depression may reduce individual fitness and therefore increase the populations extinction risks (O'Grady, Brook et al. 2006). Efficient dispersal plays an important role, in this regard, as it increase patches recolonization rate but also lower the level of inbreeding (Skelly, Werner et al. 1999). Amphibians appear to have a low connectivity thus being even more strongly affected by such events (Blaustein, Wake et al. 1994). It appears however that connectivity varies a lot between species (Smith and Green 2005). Making generalities is therefore not possible. Performing genetic studies on habitat reduction effects and population isolation individually for each species are thus major topics in conservation biology of amphibians (Amos and Balmford 2001).

Determining general state of population connectivity is however not enough. Indeed, conservation measures are most of the time restricted by their resources. They are thus forced to work on restrained units. This raises the question of the conservation unit. This is an important concept in conservation biology as conservation units are the focal units for management (Moritz 1994). Several studies have tried to assess different methods to define such units (e.g. Moritz 1994; Petranka, Smith et al. 2004). They state that conservation units should present demographic and statistical independence. In other words it is important to determine whether units act like an essentially panmictic

population or like metapopulations. Statistical independence can be tested using methods that can infer the number of statistically differentiated units present in the whole sample area. This can be done by testing for sampling units statistical differentiation (Goudet, Raymond et al. 1996), or by assigning individuals to clusters independently of their sample of origin (Pritchard, Stephens et al. 2000). Demographic independence can be inferred by estimating the number of migrant the units share per generation, i.e. the gene flow. Our estimation of the gene flow was not made through mark-recapture studies but by assuming that the gene flow is related to the F_{ST}. The number of migrant per generation can, thus, be found through the equation Ne*m = ¼ ((1/FST) – 1) (Wright 1943). This equation was inferred from island based models, however it has been shown to be robust in other models such as finite island models or stepping-stone models (Slatkin 1985). One should however keep in mind that interpretation of F_{ST} in term of gene flow are problematic in conservation biology due to the uncertainty linked to these estimates (Waples 1998).

In this study we investigated the genetic structure of fire salamander (*Salamandra salamandra*) populations. Their populations in Europe originate from two postglacial recolonisation waves (Steinfartz, Veith et al. 2000). Phylogeny of the *S. s. salamandra* and *S. s. terrestris* sub-species, which are the two sub-species present in Switzerland, remains, however, unresolved. It seems that they should be considered like two sub-populations rather than two sub-species (Steinfartz, Veith et al. 2000). In our case we are in presence of *S. s. salamandra*. This species has a complex life cycle with an aquatic larvae stage and a terrestrial post-larvae life. It is therefore, like most amphibians, linked to aquatic breeding sites. Those breeding sites are small rivers or streams with slow

current and oxygenated water. Their terrestrial habitat is most often mixed deciduous forest but the specie can also inhabit other types of forests (Catenazzi 1998).

Its connectivity is thought to be low (Smith and Green 2005) and females to be philopatric (Thieseier and Grossenbacher 2004). We can therefore state that habitat loss or fragmentation might have a negative effect on their population. Nevertheless *S. salamandra* has been classified by the IUCN as a least concern species (http://www.iucnredlist.org/) as its population are thought to be big enough and the pressure of human impact on them low. However, sightings of adult individuals in central Switzerland are decreasing suggesting that their may even so be threatened (http://www.karch.ch/).

Our study aims, are first to determine the level of structure between our sample sites, using genetic analyses of population structure, and investigate the presence of recent reduction in population size. Second it is to investigate the level which could be used as a coherent conservation unit for this specie.

Methods:

Study area and field sampling

The study was conducted in the spring 2006, in the Luzern canton, central Switzerland. The study area is composed of a main central agricultural area, the city of Luzern (GPS coordinates: 665’000-212’500) to the south-east which is the most urbanized area in the canton, lots of scattered small villages, the highway that splits the canton in two and the pre-Alps which constitute a woody area irrigated by lots of small streams (see Figure 1).

A total of 19 populations were selected, according to a distribution map of *Salamandra salamandra,* and then surveyed within the 1493 km^2 study area (see Figure 1). The map was created according to a data base of sightings provided by the KARCH. A total of 380 larvae of *S. salamandra* were sampled in our 19 sampling sites. Twenty larvae were sampled in each sampling site. The larvae were found in small rivers or streams most of the time where the current is slow and shelter available. They were caught using small nets and were collected the same day to prevent recapture. They were collected in different streams, whenever possible, and different part of those streams, on a 100-300 meters transect, to minimize the chance of collecting individuals from the same clutch. Indeed gravid females normally lay between 20 and 40 larvae in their laying site (Rebelo and Leclair 2003). One centimeter of the tail's veil was collected using sterilized scissors for each larva. Tail's veil shortening does not alter significantly their ability to swim or to escape predators (Wilbur and Semlitsch 1990; Van Buskirk and McCollum 2000). After tissue collection the larvae were released on the place where they were caught. The tissues were conserved in a 1.5ml sterile eppendorf containing 90% ethanol and stored at 5˚C before analysis.

Finally the coordinates for each sampling site were taken using Global Position System (10 m precision).

DNA extraction, PCR conditions

For about 5% of the samples, genomic DNA was extracted using the QIAgen DNeasy™ Kit (QIAgen), following the manufacturer's protocol. However PCR amplification gave poor results with a very low rate of amplification success. This may be due to the

polysaccharides present in the mucus that surrounds the tail's veil (Dubey and Fumagalli pers. comm.). It seems that extractions using the QIAgen DNeasy™ Kit (QIAgen) are not able to retain those polysaccharides in the columns. We therefore decided to change the extraction method. The genomic DNA obtained with QIAgen DNeasy™ Kit (QIAgen) extractions was purified subsequently using a phenol/chloroform protocol (Elder 1983). A CTAB method was chosen as it was designed to resolve the problem of seeds extraction, which contains a lot of polysaccharides. The other tissues were then extracted using a CTAB extraction (Lipp, Brodmann et al. 1999). The DNA was eluted in a 50-µl Tris-HCl solution and then stored at -20˚C.

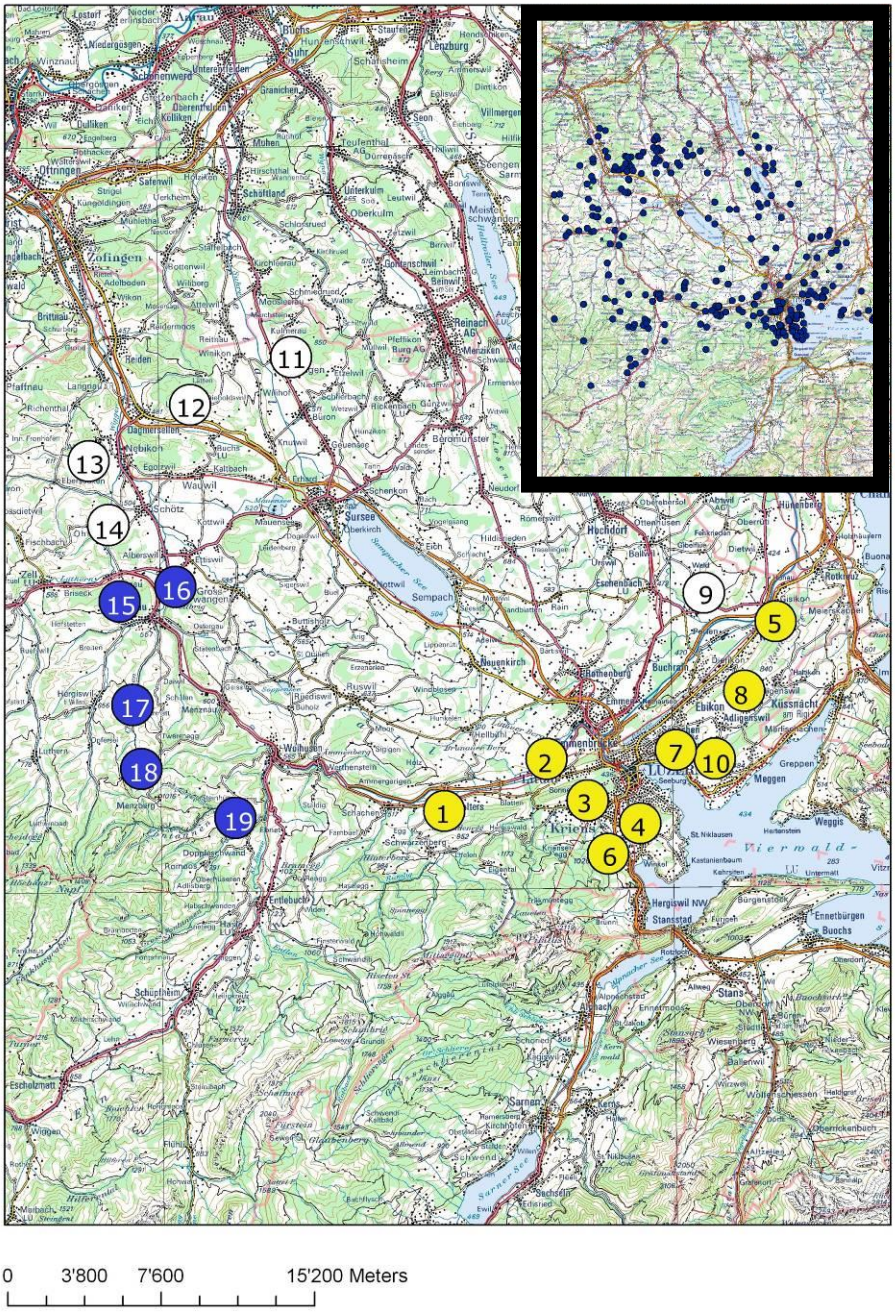

Figure 1: Map of the study area with the sampling sites (circles). The yellow circles represent sampling units that could be assigned to the first cluster, blue circles: second cluster; white circles: sampling units that could not be assigned according to Bayesian analyses (see Results). On the upper right corner we have all the sightings of adults or larvae for the same region.

Nine microsatellite loci (Sal E2, Sal E5, Sal E6, Sal E7, Sal E8, Sal E11, Sal E12, Sal 3, Sal 29 Steinfartz, Kusters et al. 2004) were amplified following the protocols described in Steinfartz, Kusters et al. (2004). PCR amplifications were carried out separately for each locus in 20 µl final volumes containing: 100-250 ng DNA, 1.5mM $MgCl_2$ (2.0 mM for Sal E2 and Sal 29), 0.5-µM of each primers (0.3-µM for Sal 29), 0.2-mM of dNTPs (0.5-mM for Sal 29), 0.2-µM Albumin (BSA) (none for Sal 29), and 0.5 U of QIAgen Taq polymerase for Sal E5, Sal E6, Sal E8, Sal E11 and Sal E12 others had 1U. All the F primers were labeled with a fluorescent dye (HEX, FAM or NED) on the 5'end.

The PCR profile for all loci was the same except for the annealing step and the number of cycles (see Table 1): (i) 95˚C for 5 min, (ii) 94˚C for 30 sec, (iii) T°C annealing for 1 min, (iv) 72˚C for 1min, (v) return to step (ii) for the fixed number of cycles, (vi) 72˚C for 10 min.

After checking amplification success on 1.5% agarose gel samples were analysed on an ABI 3100 sequencer (Applied Biosystems). Data collection, sizing of the bands and analysis were done using GeneMapper v3.7 (Applied Biosystems). Bins for each allele were created manually and afterwards each allele call was checked by eye.

Locus Description, allelic dropout, null alleles and linkage disequilibrium

The first step was to compute descriptive statistics and analysis on our loci in order to choose the ones we would keep for structure analyses.

(i) Locus description:

Descriptive statistics about our loci were obtained through analyzes of the genotypes with the FSTAT 2.9.4 program (Goudet 1995). Those values are the number of alleles, the expected and the observed heterozygsity per locus (Nei, 1987). Deviation from H-W within samples and population differentiation per locus and overall was also investigated with FSTAT 2.9.4 (Goudet 1995) through randomization based tests using the FIS statistics for H-W within samples and log-likelihood statistic *G* for population differentiation (Goudet, Raymond et al. 1996). We used for those test corrected level of significance with a 5 % nominal value.

(ii) Null alleles and allelic dropout:

Presence of null alleles or allelic dropout was tested using the program Micro-Checker 2.2.3 (Van Oosterhout, Hutchinson et al. 2004).

(iii) Genotypic linkage disequilibrium:

Overall genotypic linkage disequilibrium between each pair of locus (Goudet, Raymond et al. 1996) were tested using the FSTAT 2.9.4 program (Goudet 1995). We used for this test corrected level of significance with a 5 % nominal value.

Population structuring

(i) F-statistics:

Structuring values as well as descriptive values of our sampling units were then estimated. The number of alleles and the allelic richness per sampling unit, as well overall F_{IS} and F_{ST} (Weir and Cockerham 1984), F_{IS} per sampling unit and pairwise F_{ST} between all sampling units were computed using the FSTAT 2.9.4 program (Goudet

1995). The significance of per sampling unit F_{IS} was tested randomizing alleles within sampling units. Significance of the overall F_{IS} and F_{ST} was previously tested (see locus description). Pairwise tests of population differentiation using randomization based tested were made in order to test for the significance of each pairwise F_{ST}.

(ii) Multivariate analysis:

We first used multivariate analysis to test for sample unit clustering. We therefore used the PCAGEN 2.0 (Goudet 1999) program to compute a principal component analysis based on the genotypes of the individuals and the sampling unit they were issued.

(iii) Bayesian clustering:

In a second step, sampling unit's structure was investigate through individual assignment using a Bayesian admixture procedure implemented in STRUCTURE 2.1 (Pritchard, Stephens et al. 2000).

We assessed sampling unit's structure assuming that sampled individuals belong to an unknown number of K genetically distinct clusters. Posterior probability values for K ("Log probability of data"; L(K)) were estimated assigning a prior from one to twenty. Using only this parameter as described by Pritchard *et al*. (Pritchard, Stephens et al. 2000) it was not obvious which number of clusters (K) best fits our dataset. We thus followed the recommendation of Evanno *et al*. (Evanno, Regnaut et al. 2005) and calculated the ΔK statistic, which is based on the rate of change in the "Log probability of data" between successive K values.

Consequently, we chose the value of K showing the highest ΔK and then evaluated the individual membership coefficient (q_{ind}) to the inferred clusters. For these analyses, we used the admixture model implemented in STRUCTURE, assuming that sampled individuals belonged to K genetically distinct clusters without using any prior population information.

We performed 20 repetitions of 50,000 iterations after a burn-in period of 50,000 iterations for each K. The admixture model was used for all simulations. We then used

the population assignment value to determine more precisely which population belonged to which cluster and the ones that could not be assigned. The threshold value for population assignation was fixed at 80%.

(iv) Clusters comparison:

Considering the fact that the Bayesian clustering gave us two distinct clusters of population in our study area (see Results), clusters were compared for various sampling units factors. Those analyses have only been made on the sampling units that could efficiently be assigned to one of the clusters determined by the Bayesian analysis (see Table 4). We compared the sampling units of our two clusters on allelic richness, observed heterozygosity, expected heterozygosity, F_{ST} and F_{IS}. Those comparisons were made using the "comparison among groups of samples" applet of FSTAT 2.9.4 (Goudet 1995). This applet uses permutation-based tests.

(v) Isolation by distance:

Overall isolation by distance was investigated using the mantel test of FSTAT 2.9.4 (Goudet 1995) on a dataset composed of the pairwise $F_{ST}/(1- F_{ST})$ and the natural logarithm of Euclidean distance between sampling units for each pairs of sampling units (Rousset 1997).

To determine whether the structure between the two clusters obtained through Bayesian clustering (see results) was strictly due to isolation by distance, we run a partial mantel test using FSTAT 2.9.4 (Goudet 1995) on a restricted dataset (including only the sampling units that could significantly be assigned to a cluster) composed of the pairwise

$F_{ST}/(1- F_{ST})$, the natural logarithm of Euclidean distance and pairwise co-membership to a cluster (coded by a binomial variable, 0 for two sampling units present in the same cluster and 1 for sampling units belonging to different clusters) for each pairs of sampling units.

In a second step isolation by distance between sampling units was tested independently in each cluster.

Identification of the "smallest conservation units"

To uncover the smallest conservation unit we first run pairwise population differentiation tests based on the log-likelihood statistic G (Goudet, Raymond et al. 1996) between two units by arbitrarily fixing the "population" level. We would therefore assume that two units can be considered as distinct "populations" when they can be differentiated by a log likelihood G test (Waples and Gaggiotti 2006). We fixed two hypothetic different "population" levels. By considering the "population" level as (i) the basic sample unit which includes all the transects, ranging from 100m to 300m along a stream, done on isolated or groups of streams taken altogether and (ii) the different transects made in the streams or groups of streams within each sampling unit taken independently. The pairwise test for population differentiation of the first level was already made earlier (see F-statistics section). We then ran 19 pairwise test of population differentiation, for each sampling unit independently, between the different transects.

Second we used hierarchical F-statistics (Goudet 2005) to compute overall F_{ST} values and overall population differentiation p-values for each level. The levels here were the two previous levels (i), here the population level, and (ii), here the sub-population level, plus

the previously defined cluster level. We thus omitted for this second analyze the populations that could not be assigned to a cluster by Bayesian clustering (see Bayesian clustering and table 4). Hierarchical estimates of *F*-statistics, differentiation of sub-population within population (intra-pop, F_{SP}), of population within each cluster (intra-cluster F_{PC}) and of population between the two clusters (inter-cluster, F_{CT}) were obtained using HIERFSTAT (Goudet, 2005), implemented in R (Core Team Development). Significance of those values was tested through randomization-based tests.

Results:

Locus description, allelic dropout, null alleles and linkage disequilibrium

(i) Locus description:

The number of alleles per locus ranges from two to 12 with a total of 62 alleles overall loci. Exact number of alleles per locus is presented in Table 1. The locus Sal E5 has been conserved even though it presents only two alleles, as one of the alleles is mainly present in the western populations (frequencies in sampling units 1-10 = 0.002 and in sample units 11-19 = 0.253) supporting therefore non-negligible information. The other loci present enough alleles to be informative for our purpose (Table 1). For only one locus, Sal E12, did the expected heterozygosity diverge from the observed heterozygosity (see Table 1), suggesting that there are no important genotyping errors. As expected from the observed and expected heterozygosity Sal E12 is the only loci that significantly differs from H-W proportions (see Table 1). Analyses were consequently run on a dataset including or not this marker. As it yields similar results, we decided to keep this marker

in the data for the following analyses. The overall and locus independent tests for population differentiation (see F_{ST} values in Table 1) all gave significant results suggesting all the loci will be informative for structure analysis.

(ii) Null alleles and allelic dropout:

There was no evidence for allelic dropout in any populations for all of the loci. However in four populations we could detect null alleles for one locus (Sal E7 for pop. 6, Sal E12 for pop. 7, Sal E11 for pop. 12 and Sal E8 for pop. 15). As the presence of null alleles was not detected in all other 18 populations for each of these loci, these markers were kept in the following analyses.

(iii) Genotypic linkage disequilibrium:

Testing for overall genotypic linkage disequilibrium between loci, loci Sal E2 and Sal E6 were significantly estimated to be under linkage disequilibrium. We therefore took the locus that presented the less genetic diversity (Sal E6) off our analyses.

Table 1: Size range (in base pairs), number of individuals (*Nind*), number of alleles (*Na*), Nei's (Nei, 1987) estimation of observed heterozygosity (*Ho*) and expected heterozygosity (*Hs*), Weir and Cockerham (1984) F_{IS} and F_{ST}, annealing temperature in ˚C (*Ta °C*), the number of cycles for each primer and the fluorescent dye type (*Fd type*).

Locus	Size range (pb)	*Nind*	*Na*	*Ho*	*Hs*	F_{IS}	F_{ST}	*Ta ˚C*	*N of cycles*	*Fd type*
Sal E2	210-302	393	11	0.60	0.63	-0.009	0.051***	53	45	FAM
Sal E5	182-194	399	2	0.18	0.17	-0.059	0.209***	62	35	FAM
Sal E6	282-304	400	5	0.32	0.34	0.058	0.045***	59	40	HEX
Sal E7	154-250	400	6	0.59	0.62	0.038	0.099***	53	40	HEX
Sal E8	138-214	400	9	0.66	0.66	0.005	0.116***	58	35	NED
Sal E11	224-280	358	5	0.37	0.39	0.058	0.075***	50	40	FAM
Sal E12	162-314	400	10	0.53	0.57	0.061*	0.075***	58	35	NED
Sal 3	182-258	397	12	0.42	0.45	0.058	0.136***	53	40	NED
Sal 29	150-190	400	7	0.62	0.60	-0.038	0.095***	58	40	HEX
Overall		400	62	0.50	0.51	0.018	0.099***			

*< 0.05, ** < 0.005 and *** < 0.0005.

Population structuring

(i) F-statistics:

The number of individuals genotyped per sampling unit ranges from 17 to 20 the number of alleles per sampling units ranges from 25 to 41, the allelic richness ranges from 2.93 to 4.57 and the mean F_{ST} per sampling unit ranges from 0.07 to 0.14. For exact values per sampling unit refer to Table 2. The test for overall deviation from H-W within samples was not significant (p-value = 0.08) the F_{IS} per sampling unit were also not significant (all of them). Overall and per sampling unit F_{IS} values should therefore be taken with care and interpretation about their values avoided. The overall F_{ST} value calculated on the eight markers is 0.099 (p-value < 0.0005) suggesting a well marked differentiation between our sampling units. Mean pairwise F_{ST} values per sampling units ranges from 0.06 to 0.14 (mean = 0.096, SD = 0.019). Exact mean pairwise F_{ST} values per sampling units can be found in Table 2. The pairwise F_{ST} for all pairs of sampling units are resumed with their significance in table 3. Pairwises F_{ST} ranges from 0.02 to 0.23. Only 5% of those values are not significant indicating that our sampling units are statistically well differentiated from one another and that we can interpret the pairwise F_{ST} values.

Table 2: Coordinates, number of individuals sampled, number of alleles, allelic richness and the mean F_{ST} based on the average pair-wise F_{ST} between the focal sampling unit and all the others.

Sampling units	coordinates		number of individuals	number of alleles	allelic richness	mean F_{ST}	mean F_{IS}
	x	y					
1	656600	209495	25	28	3.31	0.09	-0.097
2	661665	212105	23	28	3.16	0.10	0.019

3	663685	210028	20	29	3.39	0.08	-0.002
4	666304	208894	17	25	3.03	0.08	0.121
5	673025	218950	20	29	3.30	0.11	0.11
6	664718	207378	20	28	3.31	0.09	0.08
7	670000	212135	18	25	2.93	0.13	-0.022
8	671460	215495	19	34	3.83	0.09	0.103
9	669450	220360	20	31	3.55	0.14	0.011
10	668080	212530	25	25	2.98	0.10	-0.019
11	644000	229711	20	36	4.04	0.09	-0.055
12	649007	232093	25	30	3.35	0.09	0.119
13	638972	226920	20	41	4.57	0.07	0.013
14	639945	223736	22	39	4.28	0.06	0.018
15	640504	219941	18	34	4.08	0.09	-0.023
16	643264	220666	20	33	3.89	0.10	0.064
17	641137	214778	20	30	3.40	0.12	-0.033
18	641572	211593	20	37	4.04	0.11	0.075
19	646250	209167	20	35	3.82	0.10	-0.057

Table 3: Pair-wise F_{ST} (below diagonal) per pair of sampling units and their significance (above diagonal); probability values are estimated over 17'000 permutations.

	pop1	pop2	pop3	pop4	pop5	pop6	pop7	pop8	pop9	pop10	pop11	pop12	pop13	pop14	pop15	pop16	pop17	pop18	pop19
pop1		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
pop2	0.02		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
pop3	0.06	0.07		NS	*	NS	NS	*	*	*	*	*	*	*	*	*	*	*	*
pop4	0.06	0.07	0.03		*	NS	*	*	*	*	*	NS	*	*	*	*	*	*	*
pop5	0.07	0.08	0.08	0.04		*	*	*	*	*	*	*	*	*	*	*	*	*	*
pop6	0.09	0.09	0.02	0.01	0.08		NS	*	*	*	*	*	*	*	*	*	*	*	*
pop7	0.08	0.09	0.06	0.07	0.11	0.06		*	*	*	*	*	*	*	*	*	*	*	*
pop8	0.10	0.08	0.04	0.02	0.06	0.03	0.06		*	*	*	*	*	*	*	*	*	*	*
pop9	0.12	0.14	0.10	0.15	0.19	0.12	0.19	0.15		*	*	*	*	*	*	*	*	*	*
pop10	0.06	0.08	0.07	0.03	0.07	0.09	0.12	0.10	0.19		*	*	*	*	*	*	*	*	*
pop11	0.08	0.08	0.09	0.10	0.12	0.09	0.13	0.09	0.14	0.10		*	*	NS	*	*	*	*	*
pop12	0.10	0.11	0.05	0.07	0.12	0.08	0.15	0.11	0.06	0.09	0.10		*	NS	*	*	*	*	*
pop13	0.08	0.10	0.08	0.06	0.07	0.09	0.11	0.08	0.15	0.06	0.06	0.09		*	*	*	*	*	*
pop14	0.07	0.06	0.04	0.07	0.08	0.06	0.11	0.05	0.10	0.07	0.02	0.05	0.04		*	*	*	*	*
pop15	0.10	0.12	0.10	0.08	0.14	0.09	0.13	0.08	0.12	0.11	0.05	0.10	0.03	0.05		*	*	*	*
pop16	0.14	0.16	0.10	0.12	0.18	0.13	0.19	0.14	0.13	0.12	0.07	0.07	0.07	0.07	0.05		*	*	*
pop17	0.16	0.18	0.16	0.16	0.19	0.17	0.23	0.18	0.19	0.14	0.07	0.13	0.06	0.08	0.07	0.03		NS	*
pop18	0.14	0.17	0.14	0.15	0.18	0.15	0.21	0.17	0.16	0.14	0.07	0.10	0.05	0.07	0.07	0.02	0.00		*
pop19	0.12	0.14	0.11	0.12	0.14	0.13	0.18	0.13	0.13	0.12	0.09	0.10	0.04	0.07	0.06	0.03	0.05	0.02	

(ii) Multivariate analysis:

The multivariate analysis performed with PCAGEN 2.0 (Goudet 1999) could only find one significant axis, the first axis explaining 37.60 % of the whole variance (p-value = 0.001). The second axis explains a fair part of the variance (14.62%) but is not significant (p-value = 0.68). We can notice in Figure 2 that two or three groups, that are consistent with geographical data (see Figure 1), emerge from visual examination of the graph. The left group composed of sampling units 15-19, a center group composed of sampling unit 9, 11, 12, 13, 14 and 15 and finally a group with sampling units 1-8 and sampling unit 10 to the right.

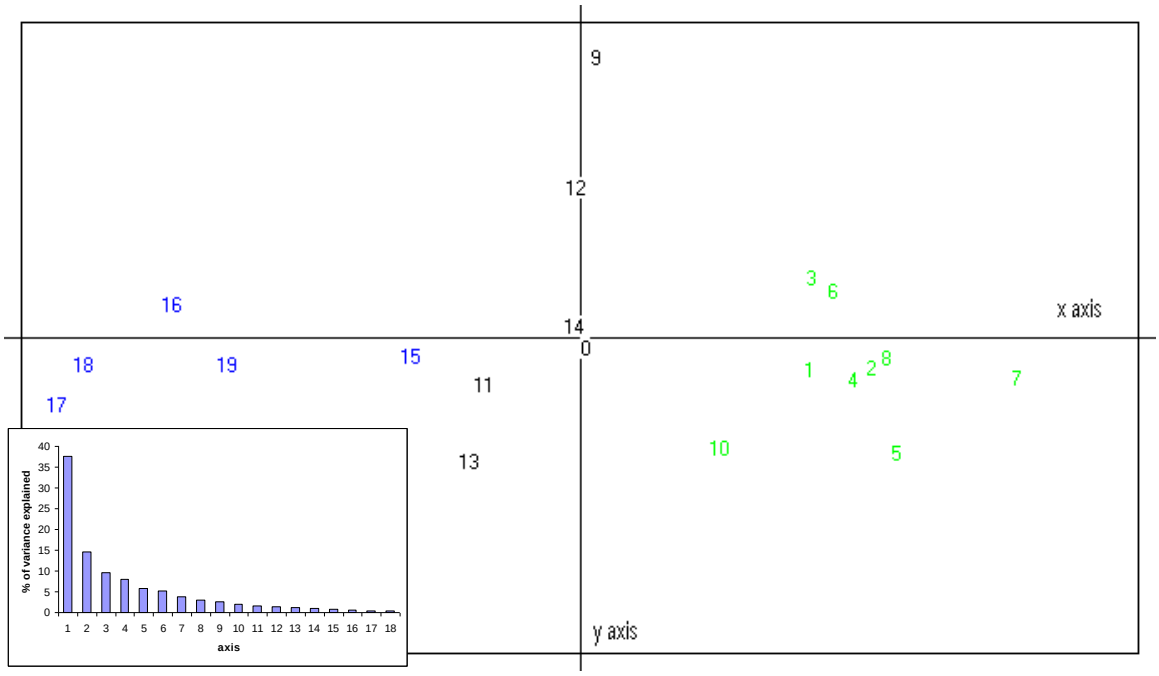

Figure 2: Graphic of the multivariate analysis with the x axis being the axis which explains most of the variance (37.6% of explained variance v, p-value = 0.001) and the y axis being the second axis which explains most of the variance (14.62% of explained variance, p-value = 0.68) and the sampling units represented by their numbers. Colors represent the two clusters given by

STRUCTURE 2.1. In the bottom left we have a graph of the percentage of variance explained by the different axis of the PCA.

(iii) Bayesian clustering:

Individual assignment by Bayesian analysis suggests that the most probable number of clusters present in our study area is two (Figure 3 and 4). Figure 3 express clearly the difficulty of directly using the mean L(K) to estimate the most probable number of clusters. One would interpret those results by assessing that the most probable number of clusters present in our population ranges from 2 to 8 as clear difference can not be made between the mean L(K) for each of those K. Evanno *et al*. (2005) method using the ΔK statistic allows, as shown in Figure 4, a more accurate interpretation of the results. One should however keep in mind that it will often induce the more conservative choice (Ref, lire Evanno et al., 2005). K=2 (see Figure 4) should therefore be taken as the uppermost level of structure in our area. Out of the 19 sampling units, 14 could be assigned to a cluster with a high posterior probability. Sampling units 1 to 8 and sampling unit 10 have a >80% probability to belong to cluster 1, sampling units 15 to 19 have a >80% probability to belong to cluster 2 and sampling units 9, 11, 12, 13 and 14 could not be assigned to any cluster (Table 4). For a geographical representation of those results, report to Figure 1. Those results are consistent with the ones given by the multivariate analysis except for sampling unit 15. We can see that in the multivariate analysis graphical results (Figure 2), would induce to accept the sampling unit 15 as member of the center group but with Bayesian clustering sampling unit 15 has a 82% probability (see Table 4), to belong to the second cluster (the blue, see Figure 1).

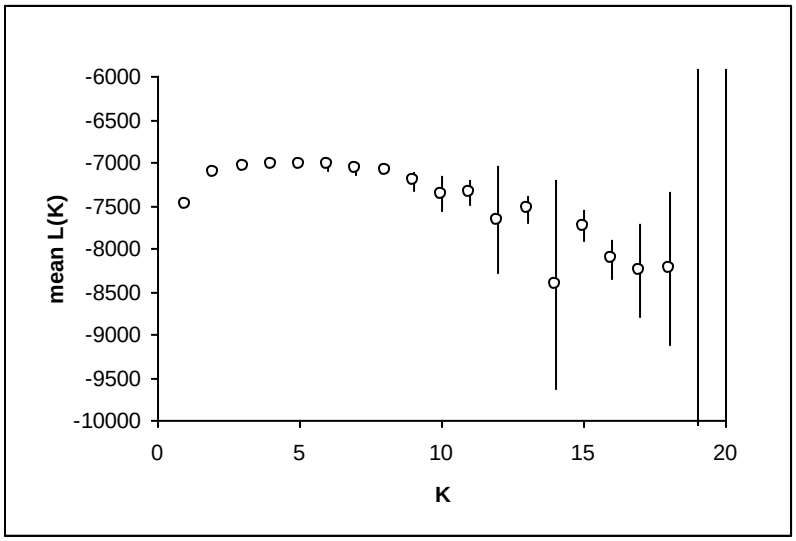

Figure 3: Graphic of the mean L(K) (±SD) over 20 runs for each probable number of clusters (K), x axis probable number of K and y axis mean –Ln(K) over 20 runs

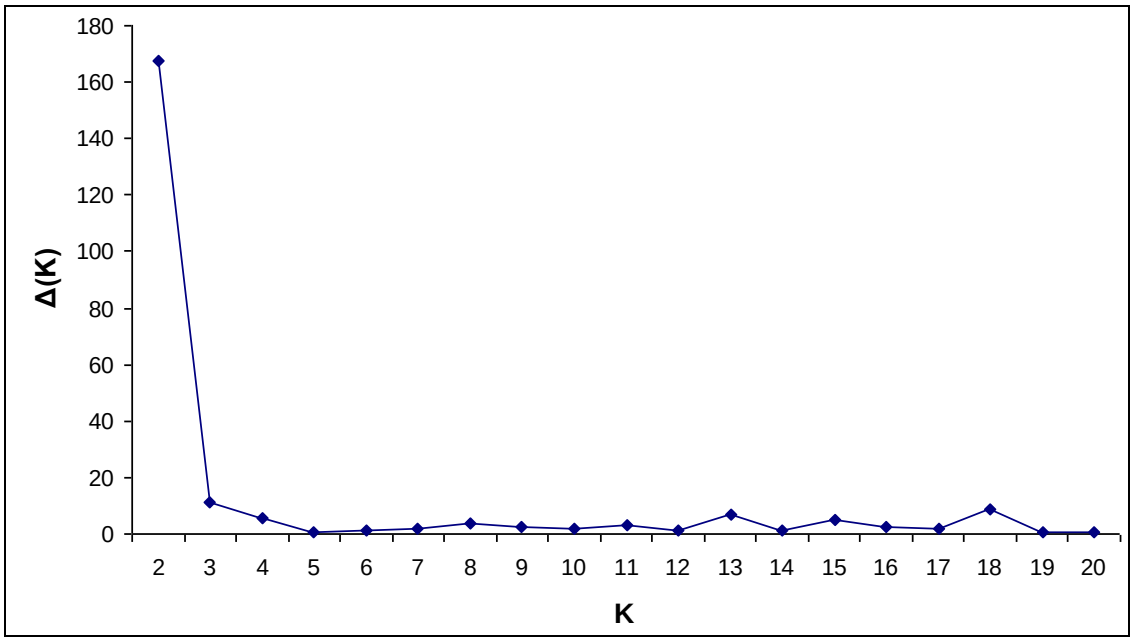

Figure 4: ΔK calculated as ΔK = m|L''(K)| / s|L(K)| in function of the number of clusters (K).

Table 4: Population assignation probability to the two clusters, CI assignation probability to the cluster I, CII assignation probability to the cluster II.

Population	CI	CII
Pop 1	**0.898**	0.102

Pop 2	**0.872**	0.128
Pop 3	**0.828**	0.172
Pop 4	**0.894**	0.106
Pop 5	**0.926**	0.074
Pop 6	**0.837**	0.163
Pop 7	**0.812**	0.188
Pop 8	**0.872**	0.128
Pop 9	0.605	0.395
Pop 10	**0.852**	0.148
Pop 11	0.404	0.596
Pop 12	0.493	0.507
Pop 13	0.396	0.604
Pop 14	0.5	0.5
Pop 15	0.178	**0.822**
Pop 16	0.112	**0.888**
Pop 17	0.063	**0.937**
Pop 18	0.057	**0.943**
Pop 19	0.126	**0.874**

(iv) Clusters comparison:

The comparison among clusters, using randomization-based tests, gave significant results for the allelic richness, the observed and the expected heterozygosity. Allelic richness, expected heterozygosity and observed heterozygosity are significantly higher in the cluster II than in the cluster I (respective p-values = 0.012, 0.001, 0.001). Still using randomization based tests the F_{ST} and F_{IS} of the two clusters do not differed significantly. Values of allelic richness, observed and expected heterozygosity for each clusters are: for cluster 1 allelic richness = 3.2, Ho = 0.45 and Hs = 0.46, for cluster 2 allelic richness = 3.8, Ho = 0.56, Hs = 0.56.

(v) Isolation by distance:

The mantel test performed on pairwise F_{ST} and the natural logarithm of Euclidean distance, following Rousset (1997), over all the pairs of populations gave significant

results with a strong correlation (r = 0.42, R^2 = 18.12, p-value < 0.001) implying strong effect of distance on the isolation of our populations (Figure 5).

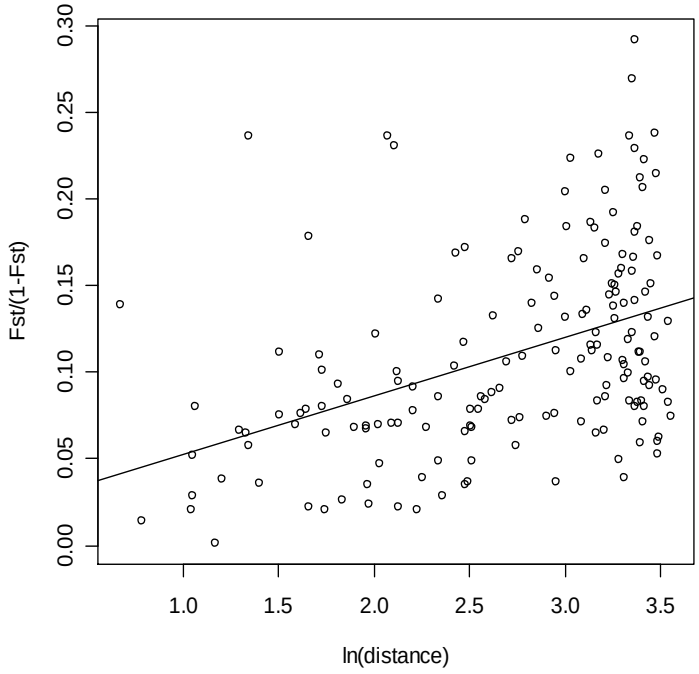

Figure 5: Relationship between pair-wise FST and pair-wise geographical distance between pairs of sampling units, following Rousset (1997).

The partial mantel test performed in order to know if the structure obtained between our two clusters was strictly due to isolation by distance gave significant results for distance (p-value < 0.001) and clustering (p-value < 0.001). Including clusters co-membership increased the explained variance compared to the model with distance alone (R^2 increases from 51.1 to 64.62 of variance explain).

No significant isolation by distance could be found between sampling units of a same cluster (cluster 1: p-value = 0.35, cluster 2: p-value = 0.45).

Identification of the "smallest conservation units"

Hierarchical F-statistics gave significant results for all of the fixed levels. As expected, the defined clusters are statistically differentiated and they yield a great amount of the structure present in our area (F_{CT} = 0.093, p-value < 0.01). The population level is also statistically differentiated within our clusters with lower but still non-negligible structure between the population units (F_{PC} = 0.045, p-value < 0.01). Surprisingly, the sub-population level also appears to be statistically differentiated, however the structure between the sub-population units is quite low (F_{SP} = 0.009, p-value < 0.01).

The population differentiation using pairwise tests based on the log-likelihood G gave significant results for our first level (i) (see methods). By fixing the smallest conservation unit at the (i) level 95 % of the units could be defined as statistically differentiated "populations" (see table 3), for the (ii) level few of the units could be distinguished as statistically differentiated. This result is somewhat in contradiction with the one obtained by hierarchical F-statistics. As expected from the hierarchical F-statistics most of the units in level (ii) should have been significantly differentiated. This may mainly be due to statistical power reduction. Indeed the pairwise analyze suffers from a low number of replicates for the (ii) level as the tests are conducted separately for the 19 sampling sites. More credit should therefore be given to the hierarchical F-statistics results (Goudet pers. comm.).

Discussion:

We found significant genetic structure (F_{ST} = 0.099***) and isolation by distance (r = 0.42, R^2 = 18.12, p-value < 0.001) among populations of *S. salamandra*, which was

expected since salamander populations are generally subdivided genetically (Larson, Wake et al. 1984; Routman 1993). This structure level represents a number of effective migrants per generation of 2.27 over a mean distance between sampling units of 18.73 km. This is somewhat contradictory as researches on adult migrations found that they tended to move on average 200-400 meters with maximum migration distances of 3156 meters over periods of one year (Catenazzi 1998). This could be explained by the fact that population connectivity in Amphibians is mostly mediated by juveniles (Madison 1997; Guerry and Hunter 2002; Rothermel 2004), implying that mark-recapture studies on adults might underestimate the real migration potential of individuals. On comparable geographical scales and in heterogeneous habitat, widespread Amphibian species present equivalent structure among populations. For example, Zamudio and Wieczorek (2007) found similar structuration for the spotted salamander (F_{ST} = 0.073), as well as Veith and al. (2002) who studied the common frog *Rana temporaria* (F_{ST} = 0.109-0.195). A study on *Plethodon cinereus* a terrestrial salamander living in continuous habitat, revealed a very low but significant fixation index (F_{ST} = 0.019; Cabe, Page et al. 2007). On the opposite, an endemic species, the streamside salamander (*Ambystoma barbouri*), presented a much greater structure (F_{ST} = 0.32). The structure between populations, in our species, though higher than for populations living in continuous habitat, spans the same range as common species population structure. This could indicate good connectivity among populations of *S. salamandra*. However, such comparisons should be taken with care because endangered species sometimes present low structure value. An example is the coastal giant salamander, (*Dicamptodon tenebrosus*) which presents relatively low

structure between its populations (F_{ST} = 0.083) even though its habitat has been vastly fragmented (Curtis and Taylor 2004).

Bayesian individual assignation analyses suggest that part of our sampling units (14 out of 19) could be clustered in two statistically coherent groups. The two clusters are not formed of the most geographically distant sampling units. For example sampling unit 1 and sampling unit 19 belong to different clusters but are separated by a distance that is smaller than some within-cluster distances. Moreover we demonstrated, using partial mantel test, that genetic differentiation between our clusters is not only due to Euclidean distances between the sampling units of the two clusters. This clustering must thus be explained by other causes. We can state different hypothesis: (i) a differential recolonization history of the region followed by maintenance of the differentiation due to biotic factors, such as migrant exclusion (ii) an abiotic barrier composed of single or cumulated landscape features.

Hierarchical F-statistics, considering the three levels: subpopulation, population and cluster, yielded an overall F_{ST} of 0.143. This value is higher than the one obtained for overall F_{ST} using FSTAT 2.9.4 (Goudet 1995). Most of this structure (F_{CT} = 0.093) is sustained by the two clusters. This level of isolation corresponds to 2.5 effective migrant per generation. The population level presents lower structure (F_{PC} = 0.045) but it is still worth considering from a demographic point of view as it corresponds to 5.5 effective migrants per generation. Sub-population structure, though significant, is very low (F_{SP} =0.001) and yields incoherent results about the number of effective migrants per generation (more than 25). These incoherent results could be expected as estimation of gene flow on the basis of F_{ST} is poorly accurate when F_{ST} is small (Whitlock and

McCauley 1999). However, we can interpret this low F_{SP} as a sign that groups of sub-population units act as panmictic populations. Sampling units and within sampling unit differentiation tests give quite the same results. Results using population differentiation tests showed that good statistical independence exists between the sampling units (population level) and ascertain the fact that group of sub-population units should be considered as panmictic populations as few units could be differentiated from each other.

The next step was to define coherent conservation units. It has been stated that they should be demographically and statistically independent (Petranka, Smith et al. 2004). From this point of view we can already exclude the sub-population level. Indeed we showed that clusters of sub-population units acted as panmictic populations. Furthermore, studies have shown that breeding sites are dynamic and strongly subject to abiotic and biotic perturbations (Petranka, Smith et al. 2004). This leaves as potential conservation unit the cluster and the population level. Both these levels appear to be demographically independent and could therefore be used as conservation units. Help about defining priorities can be found in the literature. Indeed several researchers have stated that maintaining existing forest patches as big and unfragmented as possible should be the most important issue, as individuals are unlikely to survive or travel long distances in non-forested habitats (Thompson, Gates et al. 1980; Douglas and Monroe 1981; Kleeberger and Werner 1983; Rothermel and Semlitsch 2002). However other researchers have shown that long term efficient conservation measures should take into account more than the adult core habitat (Porej, Micacchion et al. 2004). Indeed a long term vision should, at least, give the same importance to juvenile dispersion and connectivity (Carr and Fahrig 2001; Cushman 2006). With respect to the latter statement,

both levels should be considered as relevant but the cluster levels should benefit of more attention. We could therefore propose the cluster level as the upper conservation unit and the population level as the lower one. Conservation measures should set as their main goal to maintain the upper conservation unit. Conservation measures should therefore: (i) try to focus on as many smaller units as possible and (ii) pay enough attention to overall (within cluster) juvenile migration and connectivity.

No significant isolation by distance was found within the clusters. This may be a sign that there are obstacles to migration within the clusters that act unevenly on connectivity. Using least cost distance instead of single Euclidean distance in order to test for isolation by distance might be a good mean to investigate this issue. However we cannot exclude that isolation by distance was not detected because of results of a reduction of statistical power due to sample decrease.

As we stated that the clusters should be the main focus of conservation measures, we evaluated genetic diversity and overall structure of the different clusters. Comparison of the two clusters on these values revealed significant differences between them. Indeed cluster 1 present less genetic diversity and higher structure than cluster 2. Geographical investigation reveals that the cluster 1 encompasses the city of Luzern and its surroundings. Knowing that forest patches in human proximity will tend to be reduced strongly, we could state that this difference might be due to urbanisation. Indeed reduction of forest patches is known to impact Amphibian species negatively (Grialou, West et al. 2000) reducing the population size and thus genetic diversity. It could also explain higher structure as urbanisation between patches lessens connectivity. No recent bottlenecks could, however, be identified in any of the sampling units of our area.. We

can not, however exclude the fact that cluster 1 is facing perturbations due to human activities. This would somewhat go against the statement of the IUCN which implies that *Salamandra salamandra* is poorly affected by human impact. However, further examination to better understand whether historical or current factors have formed the observed pattern of population differentiation appears necessary. From other salamander species, it is well known that they survive strong alterations in their natural habitats for long periods of time (Kuzmin 1995); the long-term effects of such severe ecosystem transformations, however, remain widely unknown.

Some sampling units could not be assigned to any cluster with the Bayesian method. Exclusion, as well as strong mean pairwise F_{ST}, of the sampling unit 9 must be due to the presence, in medium to high frequency, of alleles that are rare in all other sampling units. The presence of these rare alleles could be due to the fact that this sampling unit belongs to another cluster of populations. This cluster could be composed of the unsampled populations North-East of our study area (see Figure1). Limits in those rare alleles diffusion could be explained by strong isolation from the first cluster other sampling units. Isolation may be due to landscape features acting as barriers, such as the highway or the Aare. Explanation about the other sampling units is less accurate. We can however observe that their mean pairwise F_{ST} (mean $F_{ST} = 0.080 \pm 0.027$) is slightly lower than the one of all the other sampling units taken together (mean FST = 0.097 ± 0.045) which is also true for the variance of these pairwise F_{ST}. If we compare the overall structure between each cluster and a cluster formed of our rejected sampling units (except sampling unit 9), we observe that the "rejected cluster" share same structure with the two clusters (F_{ST} with cluster 1 = 0.033**, F_{ST} with cluster 2 = 0.035**). This could imply

that those sampling units are in a transition area between the two clusters. This statement should however be taken with care. Indeed we have not sampled all the populations of the region leaving a gap of information in the central and North-East part of our study area. Collection of data about the populations in this area might, thus, bring other insights about the situation.

References:

Amos, W. and A. Balmford (2001). "When does conservation genetics matter?" Heredity **87**: 257-265.

Blaustein, A. R., D. B. Wake, et al. (1994). "Amphibian Declines - Judging Stability, Persistence, and Susceptibility of Populations to Local and Global Extinctions." Conservation Biology **8**(1): 60-71.

Cabe, P. R., R. B. Page, et al. (2007). "Fine-scale population differentiation and gene flow in a terrestrial salamander (Plethodon cinereus) living in continuous habitat." Heredity **98**(1): 53-60.

Carr, L. W. and L. Fahrig (2001). "Effect of road traffic on two amphibian species of differing vagility." Conservation Biology **15**(4): 1071-1078.

Curtis, J. M. R. and E. B. Taylor (2004). "The genetic structure of coastal giant salamanders (Dicamptodon tenebrosus) in a managed forest." Biological Conservation **115**(1): 45-54.

Cushman, S. A. (2006). "Effects of habitat loss and fragmentation on amphibians: A review and prospectus." Biological Conservation **128**(2): 231-240.

Douglas, M. E. and B. L. Monroe (1981). "A Comparative-Study of Topographical Orientation in Ambystoma (Amphibia, Caudata)." Copeia(2): 460-463.

Drost, C. A. and G. M. Fellers (1996). "Collapse of a regional frog fauna in the Yosemite area of the California Sierra Nevada, USA." Conservation Biology **10**(2): 414-425.

Elder, R. T. (1983). "Molecular-Cloning - a Laboratory Manual - Maniatis,T, Pritsch,Er, Sambrook,J." Bioscience **33**(11): 721-722.

Evanno, G., S. Regnaut, et al. (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." Molecular Ecology **14**(8): 2611-2620.

Funk, W. C., A. E. Greene, et al. (2005). "High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation." Biology Letters **1**(1): 13-16.

Goudet, J. (1995). "FSTAT (Version 1.2): A computer program to calculate F-statistics." Journal of Heredity **86**(6): 485-486.

Goudet, J. (2005). "HIERFSTAT, a package for R to compute and test hierarchical F-statistics." Molecular Ecology Notes **5**(1): 184-186.

Goudet, J., M. Raymond, et al. (1996). "Testing differentiation in diploid populations." Genetics **144**(4): 1933-1940.

Grialou, J. A., S. D. West, et al. (2000). "The effects of forest clearcut harvesting and thinning on terrestrial salamanders." Journal of Wildlife Management **64**(1): 105-113.

Guerry, A. D. and M. L. Hunter (2002). "Amphibian distributions in a landscape of forests and agriculture: An examination of landscape composition and configuration." Conservation Biology **16**(3): 745-754.

Kleeberger, S. R. and J. K. Werner (1983). "Post-Breeding Migration and Summer Movement of Ambystoma-Maculatum." Journal of Herpetology **17**(2): 176-177.

KUZMIN, S. L. (1995): The clawed salamanders of Asia. Neue Brehm-Bücherei, vol. 622. Westarp Wissenschaften, Magdeburg)

Larson, A., D. B. Wake, et al. (1984). "Measuring Gene Flow among Populations Having High-Levels of Genetic Fragmentation." Genetics **106**(2): 293-308.

Lipp, M., P. Brodmann, et al. (1999). "IUPAC collaborative trial study of a method to detect genetically modified soy beans and maize in dried powder." Journal of Aoac International **82**(4): 923-928.

Madison, D. M. (1997). "The emigration of radio-implanted spotted salamanders, Ambystoma maculatum." Journal of Herpetology **31**(4): 542-551.

Moritz, C. (1994). "Defining Evolutionarily-Significant-Units for Conservation." Trends in Ecology & Evolution **9**(10): 373-375.

O'Grady, J. J., B. W. Brook, et al. (2006). "Realistic levels of inbreeding depression strongly affect extinction risk in wild populations." Biological Conservation **133**(1): 42-51.

Petranka, J. W., C. K. Smith, et al. (2004). "Identifying the minimal demographic unit for monitoring pond-breeding amphibians." Ecological Applications **14**(4): 1065-1078.

Porej, D., M. Micacchion, et al. (2004). "Core terrestrial habitat for conservation of local populations of salamanders and wood frogs in agricultural landscapes." Biological Conservation **120**(3): 399-409.

Pritchard, J. K., M. Stephens, et al. (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.

Rebelo, R. and M. H. Leclair (2003). "Differences in size at birth and brood size among Portuguese populations of the fire salamander, Salamandra salamandra." Herpetological Journal **13**(4): 179-187.

Rothermel, B. B. (2004). "Migratory success of juveniles: A potential constraint on connectivity for pond-breeding amphibians." Ecological Applications **14**(5): 1535-1546.

Rothermel, B. B. and R. D. Semlitsch (2002). "An experimental investigation of landscape resistance of forest versus old-field habitats to emigrating juvenile amphibians." Conservation Biology **16**(5): 1324-1332.

Rousset, F. (1997). "Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance." Genetics **145**(4): 1219-1228.

Routman, E. (1993). "Population-Structure and Genetic Diversity of Metamorphic and Paedomorphic Populations of the Tiger Salamander, Ambystoma-Tigrinum." Journal of Evolutionary Biology **6**(3): 329-357.

Sarkar, S. (1996). "Ecological theory and anuran declines." Bioscience **46**(3): 199-207.

Skelly, D. K., E. E. Werner, et al. (1999). "Long-term distributional dynamics of a Michigan amphibian assemblage." Ecology **80**(7): 2326-2337.

Slatkin, M. (1985). "Gene Flow in Natural-Populations." Annual Review of Ecology and Systematics **16**: 393-430.

Smith, M. A. and D. M. Green (2005). "Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations?" Ecography **28**(1): 110-128.

Steinfartz, S., D. Kusters, et al. (2004). "Isolation and characterization of polymorphic tetranucleotide microsatellite loci in the Fire salamander Salamandra salamandra (Amphibia : Caudata)." Molecular Ecology Notes **4**(4): 626-628.

Steinfartz, S., M. Veith, et al. (2000). "Mitochondrial sequence analysis of Salamandra taxa suggests old splits of major lineages and postglacial recolonizations of Central

Europe from distinct source populations of Salamandra salamandra." Molecular Ecology **9**(4): 397-410.

Stuart, S. N., J. S. Chanson, et al. (2004). "Status and trends of amphibian declines and extinctions worldwide." Science **306**(5702): 1783-1786.

Thompson, E. L., J. E. Gates, et al. (1980). "Distribution and Breeding Habitat Selection of the Jefferson Salamander, Ambystoma-Jeffersonianum, in Maryland." Journal of Herpetology **14**(2): 113-120.

Van Buskirk, J. and S. A. McCollum (2000). "Influence of tail shape on tadpole swimming performance." Journal of Experimental Biology **203**(14): 2149-2158.

Van Oosterhout, C., W. F. Hutchinson, et al. (2004). "MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data." Molecular Ecology Notes **4**(3): 535-538.

Veith, M., M. Vences, et al. (2002). "Genetic differentiation and population structure within Spanish common frogs (Rana temporaria complex; Ranidae, Amphibia)." Folia Zoologica **51**(4): 307-318.

Waples, R. S. (1998). "Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species." Journal of Heredity **89**(5): 438-450.

Waples, R. S. and O. Gaggiotti (2006). "What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity." Molecular Ecology **15**(6): 1419-1439.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-Statistics for the Analysis of Population-Structure." Evolution **38**(6): 1358-1370.

Whitlock, M. C. and D. E. McCauley (1999). "Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1)." Heredity **82**: 117-125.

Wilbur, H. M. and R. D. Semlitsch (1990). "Ecological Consequences of Tail Injury in Rana Tadpoles." Copeia(1): 18-24.

Wright, S. (1943). "Isolation by distance." Genetics **28**(2): 114-138.

Zamudio, K. R. and A. M. Wieczorek (2007). "Fine-scale spatial genetic structure and dispersal among spotted salamander (Ambystoma maculatum) breeding populations." Molecular Ecology **16**(2): 257-274.