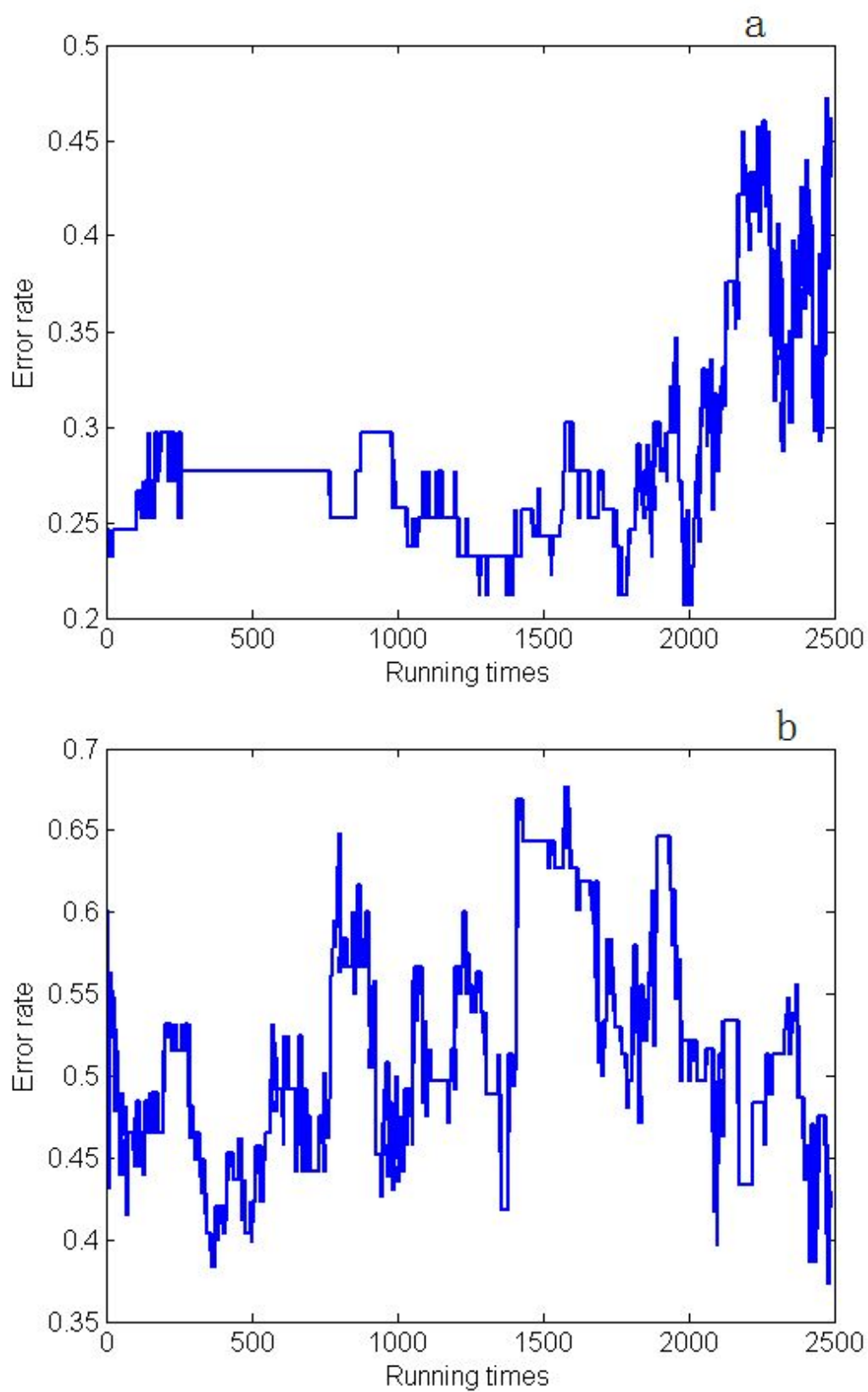


**Figure S1** Learning curves of motif finding step of PMF algorithm for dental plaques samples with SVM (a) and saliva samples with NNA (b)



**Table S1:** Microbiota signatures ('KCTCWT') in 3-class problem in pneumonia samples

Name	Id [2]	Relate to pneumonia
Actinomyces	GU737566	*
Pseudonocardia	GU737576	[21]
Mycobacterium	GU737577	*
Parvimonas	GU737581	*
Eubacterium	GU737582	[22]
Acinetobacter	GU737623	*
Aggregatibacter	HQ914699	*
Atopobium	HQ914700	*
Olsenella	HQ914714	[23]
Amaricoccus	HQ914728	
Anoxybacillus	HQ914730	[24]
Brevibacillus	HQ914731	
Enhydrobacter	HQ914734	[25]
Ferrimonas	HQ914736	
Serratia	HQ914745	[26]
Thermus	HQ914747	[27]
Peptococcus	HQ914755	[28]
Aeromonas	HQ914760	[29]
Brevundimona	HQ914762	[30]

\* top 20 genera suspiciously contributing to pneumonia [2]

**Table S2:** Microbiota signatures ('WTCGTC' ) in 2-class problem in pneumonia samples

Name	Id [2]	Relate to pneumonia
Aggregatibacter	HQ914699	*
Atopobium	HQ914700	*
Olsenella	HQ914714	*
Brevibacillus	HQ914731	
Chryseobacterium	HQ914732	[30]
Brevundimonas	HQ914762	[29]

\* top 20 genera suspiciously contributing to pneumonia [2]

**Table S3:** Classification results of dental decay data from decay plaques samples

Method	Error rate		Dimension
	On training data	On test data	
SVM/FMS	0.3385	0.4444	13
SVM/PMF	0.2876	0.2308	402
SVM/KruskalWallis	0.2319	0.3654	2705
SVM/InformationGain	0.2121	0.5192	132
SVM/Chi-square statistic	0.2462	0.3654	2997

**Table S4:** Classification results of dental decay data from saliva samples

Method	Error rate		Dimension
	On training data	On test data	

	data		
SVM/FMS	0.2583	0.4118	46
NNA/PMF	0.42	0.1169	40
NNA/KruskalWallis	0.3617	0. 3766	97
SVM/InformationGain	0.38	0.4221	2705
SVM/Chi-square statistic	0.2567	0.3961	122